Etude des méthodes de dissimulation informées de données appliquées aux supports multimédias

Sofiane Braci

N° D'ORDRE

Faculté des sciences d'Orsay

UNIVERSITÉ PARIS-SUD 11

THÈSE DE DOCTORAT

SPECIALITE : PHYSIQUE

Ecole Doctorale « Sciences et Technologies de l'Information, des Télécommunications et des Systèmes »

Présentée par :

Sofiane BRACI

Sujet :

Étude des méthodes de dissimulation informées de données appliquées aux supports multimédias

Manuscrit de thèse provisoire. Version du 13 avril 2011.

Soutenue le ..... devant les membres du jury :
M. Rémy BOYER (Maitre de conférences, L2S, Paris-Sud-11) : Directeur de thèse
M. Claude DELPHA (Maitre de conférences, L2S, Paris-Sud-11) : Co-Directeur de thèse
Mme. Françoise PRETEUX (Professeur, Ecole des Mines, Paris) : Rapporteur
M. William PUECH (Professeur, LIRMM, Montpellier) : Rapporteur
M. Azzedine BEGHDADI (Professeur, L2TI, Paris-13) : Examinateur
M. Pierre DUHAMEL (Directeur de recherche, L2S, Paris-sud-11) : Examinateur

Table des matières

Table des figures

Notations

Sauf indication contraire dans les paragraphes, nous indiquons ci-dessous les principales notations utilisées dans le document.

$\mathbf{v}$	vecteur colonne.
V	variable aléatoire.
v	variable scalaire.
$\mathbb{E}\{\cdot\}$	espérance mathématique.
$\ll$	opérateur décalage binaire de 1 bit vers la gauche.
$\gg$	opérateur décalage binaire de 1 bit vers la droite.
$\cup$	opérateur OR (ou logique).
$D(.\|\|.)$	entropie relative ou distance de Kullback-Leibler.
p_S	fonction densité de probabilité de la variable aléatoire S.
$H(.)$	entropie de Shannon.
$I(.,.)$	information mutuelle de deux variable aléatoire.

Acronymes et Abréviations

Data-hiding *Dissimulation de l'information*
AWGN *Additive White Gaussian Noise*
SS *Spread Spectrum*
SCS *Scalar Costa Scheme*
TCQ *Trellis Coded Quantization*
ST *Spread Transform*
TFA *Temporal Frame Averaging*
w.n.r. *wattermark to noise ratio*
d.w.r. *document to watermark ratio*
p.d.f. *probability density function*

Chapitre 1

Introduction

L'avènement d'internet à partir des années 90, a ouvert de larges perspectives à différentes applications dans plusieurs secteurs d'activité, tels que : la télé-éducation, la santé, les activités militaires et bien d'autres. La quantité de données numériques échangées a ainsi connu une augmentation exponentielle, facilitée par l'efficacité des réseaux de communication et l'augmentation constante des débits de transfert de données. La démocratisation des services en ligne et la disponibilité d'ordinateurs à bas coût, ont permis l'émergence de nouvelles offres de stockage et de distribution de données multimédias, générant des bénéfices importants pour l'industrie multimédia. Cependant, la gratuité d'accès aux données a facilité et banalisé le piratage des oeuvres multimédias, grâce notamment aux réseaux peer-to-peer et le téléchargement direct permettant le partage sans contrôle des fichiers audio et vidéo. Ceci a permis le développement de nouvelles techniques de dissimulation de l'information dans les œuvres dématérialisées, afin de garantir leurs traçabilité. Parmi ces techniques de dissimulation de l'information figurent le tatouage numérique destiné à l'anti-piratage et la stéganographie pour l'échange d'informations cachées.

Généralités

La dissimulation de l'information existe depuis plusieurs milliers d'années. Comme il a été décrit dans [3], rendre illisible un contenu en procédant à un cryptage n'est pas toujours la solution la plus adéquate en pratique. Dans certains environnements hostiles, cacher l'existence d'une communication est nécessaire pour éviter un certain nombre d'attaque de la part d'adversaires mal-intentionnés. Par exemple, la

stéganographie qui fait partie de la dissimulation de l'information existait déjà en Grèce Antique. L'histoire de la dissimulation de l'information a souvent été liée aux applications militaires et au contre-espionage. Durant les dernières décennies, la dissimulation de l'information a connu une grande révolution avec l'avènement d'Intenet et des supports numériques. Ainsi, la communauté a vu la publication de centaines d'articles et de communication traitant des différentes techniques d'utilisation de la dissimulation de l'information [4]. Elle est devenue l'un des outils les plus efficace dans la lutte contre le piratage multimedia et dans l'encadrement de la diffusion des données numériques.

De nos jours, on considère que le data-hiding (dissimulation de l'information) est un domaine assez mûr avec plusieurs classifications et partitionnements. Nous allons nous interesser dans ce travail à deux types de classifications : suivant l'application et suivant le modèle de communication adoptés par le système du data-hiding. Nous présentons dans ce rapport les 3 principales applications du data-hiding.

Stéganographie : elle consiste en l'altération d'un contenu d'une manière indétectable dans le but d'insérer un message (voir [5] pour plus de détails). Elle est la plus ancienne application du data-hiding, puisqu'elle est retrouvée dans l'Antiquité avec l'histoire de Herodotus [6]. Elle sert généralement à communiquer secrètement à travers des réseaux publics. Ainsi, elle est souvent utilisée dans le domaine militaire ou le contre-espionnage.

Tatouage numérique : C'est l'application la plus connue du data-hiding. D'ailleurs, le tatouage numérique est souvent utilisé comme le terme générique qui désigne tout les systèmes de data-hiding. Il est défini comme l'ensemble des modifications robustes et invisibles dans un document hôte. Ces modifications permettent d'insérer une information qui concerne directement le document tatoué, copyright par exemple, ou indirectement, tel que les droits de l'utilisateur sur le contenu (pour plus d'informations voir par exemple les références suivantes [7] [8] [9]).

Estampillage (Fingerprinting) : [10] C'est une technologie donnant la possibilité à la justice d'identifier et punir les personnes responsables de la diffusion des copies non-autorisées [10]. Pour pouvoir identifier l'utilisateur qui est à l'origine d'une distribution illicite d'une vidéo, une marque spécifique est insérée pour chaque utilisateur.

Le système doit être parfaitement sécurisé et doit avoir un bon niveau de robustesse face aux traitements licites, puisque l'empreinte (fingerprint) doit survivre même

dans les environnements les plus hostiles, tel que le réseau internet.
A noter que dans ce travail, le fingerprinting désigne uniquement la partie traçage des documents multimédias. Car la partie applicative de la thèse s'inscrit dans des projets de traçabilité des documents numériques. Nous avons jugé préférable de garder cette nomenclature sans traiter en détail les caractéristiques d'un fingerprint comme la propriété d'anti-collusion.

Il existe d'autres applications ou sous-applications, des systèmes de data hiding tel que l'optimisation des systèmes de compression ou l'authentifiation [11] [12].

Les systèmes de data hiding sont géneralement composés d'encodeur et de décodeur, l'information est transmise en utilisant un support hôte qui joue le rôle d'un canal de transmission. Ainsi, un système de data hiding est bien un système de communication [5]. C'est la raison pour laquelle il existe une deuxième classification selon le modèle de communication adopté.

Système de data-hiding sans information adjacente ou non-informé : Lorsqu'un système de data-hiding insert une information sans tenir compte des caractéristiques du signal hôte, le système est dit non-informé. Ce type de système a été initialement adopté par les premiers schémas de data hiding.
Étalement de spectre (SS ou Spread spectrum watermarking) L'étalement du spectre est à la base une technique du domaine des télécommunications numériques, elle s'est avérée très bénéfique pour le tatouage numérique en particulier pour le tatouage vidéo où l'espace d'insertion est très large. Le but principal de cette méthode est de créer la redondance pour augmenter la robustesse du système. Cette technique a été adoptée par Franck Hartung [13] pour un système de tatouage déstiné à la protection video.
Cependant, lorsque le tatouage est modélisé par un canal gaussien AWGN, le signal hôte est considéré comme un bruit ajouté à la marque en plus des dégradations subies par le document tatoué. Ceci limite considérablement les perfomances du système de marquage.

Système de data-hiding avec information adjacente (systèmes informés) : D'après les travaux de Costa [14] dans le domaine des télécommunications, le signal hôte n'est pas considéré comme un bruit ajouté dans le cas des systèmes à informa-

tion adjacente, mais un moyen d'adapter l'encodage au canal de transmission. De cette manière, le recouvrement de la marque est plus efficace et le système devient plus robuste.
Ce travail utilise principalement ce type de systèmes de tatouage et plus particulièrement les schémas suivants :

- *Quantization Index Modulation (QIM)* ou quantification par modulation de l'index, a été proposée par Chen et Wornel [15]. Elle désigne la technique de tatouage qui module d'abord l'index ou la séquence d'index avec l'information à transmettre, puis procéde à la quantification du signal hôte au quantificateur. Cette technique est très utilisée grâce à sa simplicité et à son efficacité.

- *Scalar Costa Scheme (SCS)* ou le schéma scalaire de Costa découle directement du schéma de codage proposé par Costa dans un article " Writing on Dirty Paper " [14]. Initialement, ce schéma est théorique impossible à réaliser en pratique à cause des dictionnaires de tailles infinies nécessaires pour sa mise en œuvre. Eggers et al. [1] ont proposé une implémentation sous-optimale de ce schéma, dans laquelle ils utilisent des quantificateurs scalaires pour construire les dictionnaires de Costa (d'où Scalar dans le nom donné à la méthode). Le décodage utilise un critère de décision calculé à partir du signal reçu pour estimer le message inséré.

- *Treillis Coded Quantization (TCQ)* ou quantification codée par treillis utilise un treillis associé à un dictionnaire structuré. Elle permet de réduire la complexité du système de tatouage tout en réduisant la distorsion. Pour insérer le tatouage, les chemins de treillis sont forcés par les valeurs des bits du message et les échantillons du signal hôte sont quantifiés à l'aide des dictionnaires correspondants au chemin emprunté, ce qui donnera un débit de 1 bit par échantillon. Le tatouage de chaque échantillon hôte dépend donc de l'état précédent du treillis et du symbole d'entrée. Afin d'améliorer les performances de ce schéma, l'erreur de quantification est calculée puis multipliée par le facteur de Costa et ajoutée au signal hôte. Autrement dit, nous procéderons de la même manière que le SCS sauf que le quantificateur scalaire est remplacé par un quantificateur TCQ. Au décodage, le signal reçu est re-quantifié à l'aide de l'algorithme de Viterbi pour retrouver le meilleur chemin parmi tous les

chemins du treillis possibles, les transitions de ce chemin permettront de récupérer le message inséré. Il est à noter que le tatouage TCQ permet d'avoir une meilleure robustesse que le SCS pour des rapports watermark à bruit important.

Ce manuscrit de thèse est divisé en trois grands chapitres. Le premier est un état de l'art sur la dissimulation de l'information. Il justifie une partie de nos choix concernant les systèmes de data hiding à étudier et les analyses effectuées durant cette thèse. Le deuxième chapitre traite des résultats obtenus pour des travaux autours de la stéganographie active et passive. Il donne les formulations théoriques que nous avons développé dans ce contexte ainsi que les résultats des simulations. Enfin, le troisième chapitre regroupe les études et résultats obtenus pour les systèmes du data hiding informés dans le contexte du tatouage numérique. Nous développons dans ce chapitre une étude détaillée sur les propriètés des systèmes informés dans un contexte du watermarking, en particulier la partie concernant la sécurité. Nous terminons ce document par une conclusion générale et quelques perspectives pour les travaux à venir. D'ailleurs, l'une des perspectives futures de ce travail est détaillée dans la première annexe, dans laquelle nous proposons une méthode de traçage des contenus vidéo compressés en format H.264. Notons que plusieurs parties de ce manuscrit de thèse ont fait l'objet de publications scientifiques listées dans l'annexe C.

Chapitre 2

Etat de l'art

Ce chapitre présente l'état de l'art sur la dissimulation d'information. Après quelques généralités sur les systèmes de data hiding, nous donnons les raisons qui ont motivé nos choix des systèmes informés basés sur la quantification. Considérons les trois situations suivantes :

- Des collégiens faisant passer entre eux un mot écrit sur un bout de papier en cachette de leur professeur.
- Une caissière vérifiant, à l'aide d'une lampe UV, l'authenticité d'un billet de 500 euros.
- La présence de codes barre sur toutes les canettes d'une grande marque de soda.

A priori, les trois situations présentées n'ont aucune relation entre elles. Cependant, elles ont toutes un point commun qui est celui de représenter une étape ou une partie d'un processus appartenant à la famille des schémas de *dissimulation de l'information* ou plus connu sous la dénomination anglophone *Data Hiding* ou encore *Information Hiding*.

En se basant sur les définitions de Cox et al.[5], le data-hiding est un terme général regroupant un large nombre de problèmes concernant l'insertion d'un message ou une information dans un contenu. Ainsi, dans la première situation, l'information à cacher est le mot écrit sur le bout de papier afin qu'il puisse être échangé discrètement sans que le professeur ne s'en rende compte. Ce cas peut être considéré comme une opération de stéganographie. La deuxième situation concerne la vérification d'un billet de banque en lisant *le filigrane* ou *le Watermark*. L'information dans ce cas est le filigrane alors que le support est le billet de banque. Le filigrane ou le Watermark

a donné son nom à l'une des plus importantes et des plus connues des disciplines du data-hiding. Le filigrane, plus connu dans la communauté sous le nom tatouage, est l'opération d'insertion d'une information généralement imperceptible qui concerne directement le support ou le contenu dit ″marqué″. Enfin, le dernier cas est une illustration du data-hiding pour le traçage, puisque le code barre contient, une information extraite à l'aide du lecteur à diode ou laser. Cette information contient l'identifiant du produit permetant de tracer et de remonter la chaine jusqu'au fabricant. La partie qui traite du traçage des contenus dans le data-hiding est appelée le *Fingerprinting*.

2.1 Généralités

La dissimulation de l'information traite du problème de transmission d'un signal représentant souvent un message ou une donnée à transmettre en utilisant comme support ou canal de communication un contenu hôte. Ce dernier peut prendre plusieurs formes, numérique : vidéos et images, ou encore les lettres d'un texte anodin utilisé par les services secrets dès deuxième guerre mondiale.

Définition du data-hiding

La dissimulation de l'information consiste à transmettre un message $\mathbf{m}$ via un document hôte donné par le vecteur $\mathbf{s} \in \mathcal{S}^n$ où n est un entier naturel. Le signal $\mathbf{s}$ est modifié de façon à permettre au décodeur d'extraire le message sans que les caractéristiques du signal hôte ne soient modifiées, autrement dit, l'insertion de l'information doit idéalement être invisible du point de vue statistique et perceptuel. Le schéma du data-hiding peut être résumé dans le schéma de Fig.2.1. Le message $\mathbf{m}$ est choisit dans l'alphabet $\mathcal{M}^n$. Le signal : $\mathbf{x} = \mathbf{s} + \mathbf{w}$ transmis sur le canal peut subir divers dégradations dues au canal de transmission. Pour un signal reçu donné, probablement dégradé, le décodeur donnera une estimation $\hat{\mathbf{m}}$ du message $\mathbf{m}$. D'une manière plus simple, on pourrait dire qu'un système de dissimulation de l'information consiste en une fonction de codage et une autre fonction de décodage. L'insertion de l'index $\mathbf{m} \in \mathcal{M}^n$ dans le signal hôte $\mathbf{s} \in \mathcal{S}^n$ reviendrait à transformer $\mathbf{m}$ et $\mathbf{s}$ en un autre signal $\mathbf{x} \in \mathcal{X}^n$. La clef $\mathbf{k} \in \mathcal{K}^n$ (voir Fig.2.1) est utilisée afin de rendre le signal de tatouage aléatoire pour un utilisateur non-autorisé. Elle sert à crypter notre message pour le rendre illisible à une personne non-autorisée. L'opération de

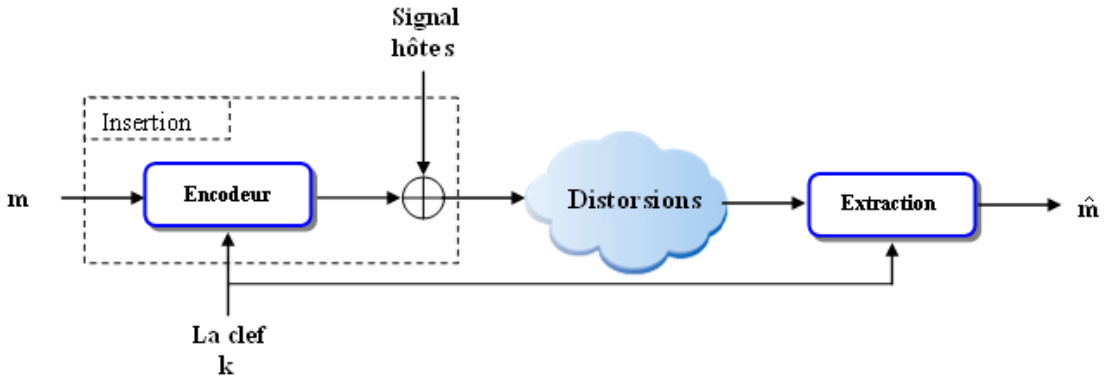


FIGURE 2.1 – Schéma général d'un système de dissimulation d'information.

chiffrement est effectuée en se basant sur le principe de Kerckhoffs " *Toute méthode de chiffrement est connue de l'ennemi, la sécurité du système ne dépend que du choix des clefs* "". La clef utilisée pour l'insertion à l'encodeur doit evidement être connue du décodeur.
Le décodeur reçoit un signal $\mathbf{y}$ correspond à la somme du signal composite $\mathbf{x}$ et un certain vecteur de bruit $\mathbf{v}$, ce dernier vient des perturbations rencontrées par le signal $\mathbf{x}$ lors de sa transmission à travers le canal. Le signal original est considéré inconnu par le récepteur (système aveugle), ce qui est le cas pour la plupart des systèmes de data hiding sont des systèmes aveugles. D'ailleurs, tous les systèmes étudiés et analysés dans ce rapport sont aveugles. Le rôle du décodeur est d'extraire le message $\mathbf{m}$ du signal $\mathbf{y} \in \mathcal{Y}^n$.
Le développement des systèmes de data hiding a permis l'apparition de plusieurs méthodes de construction du signal d'insertion et divers techniques pour l'extraction de l'information. Ainsi, chaque technique est définie par le compromis entre ces différentes carctéristiques.

2.1.1 Caractéristiques d'un système de dara-hiding

Dans Tous les schémas de dissimulation de l'information, il faut respecter un compromis entre leurs principales caractéristiques : Robustesse, capacité, invisibilité et sécurité. Ce compromis répond à certaines conditions dictées par un ensemble d'éléments : Le niveau de menace qui pèse sur l'information insérée, l'environnement dans lequel évolue le document marqué,.. etc. Ces conditions sont liées aux

types d'applications souhaités du système. Nous définissons dans la suite les principales caractéristiques principales d'un système de data hiding considérées dans la bibliographie (voir les tutoriels [16] [17]).

Capacité

C'est le nombre moyen de bits transmissible par échantillon hôte. C'est également la taille du message qu'il est possible d'insérer sans erreur sur une longueur donnée du signal hôte pour une puissance d'attaque donnée. La capacité dépend principalement de la nature du signal hôte et du compromis robustesse/capacité. Dans certains systèmes, la résistance au bruit est obtenue grâce à la redondance des bits d'informations, tel que le tatouage par étalement : En augumentant la redondance, la capacité diminue pour un gain en robustesse. Pour notre travail, la capacité du système de marquage n'est pas très importante puisque un identifiant comportera au maximum de 50 bits d'informations. Ce qui est faible relativement au support dans lequel l'identifiant est inséré.

Robustesse

Ce terme est généralement réservé au cas spécifique de la résistance aux traitements légaux : compression, recadrage, changement du contraste...etc, où l'attaque altère ou élimine la marque. Dans la protection de propriété intellectuelle et le traçage de document illicites, la robustesse prend une importance élevée puisque dans ce cas la marque doit pouvoir survivre aux différentes agressions ou toute modification du signal marqué susceptible d'éliminer le tatouage sans qu'elle découle forcément d'une mauvaise intention de la part de l'utilisateur, par exemple le changement du contraste des séquences d'images ou la compression. La protection de notre document repose donc entièrement sur le tatouage. Pour cette thèse, cette caractéristique est nécessaire pour garantir une diffusion sécurisée dans le cas du tatouage robuste et contrer ainsi les attaques du gardien actif en stéganographie.

Invisibilité

Toute dissimulation d'information engendre inévitablement des modifications sur le document hôte. Ceci impacte les caractéristiques peceptuelles et statistiques du signal original. Ainsi, le but de la majorité des systèmes de data hiding est d'impacter au minimum les cactéristiques du signal original.

Fidélité ou invibilité perceptuelle

La distorsion due au marquage du signal hôte ne doit pas dépasser le seuil de perception de l'utilisateur, un nombre important de travaux ont été réalisés sur la modélisation des système sensoriels humain [18] [19]. Les contraintes liées à l'effet de la perception des déformations engendrées par l'insertion du message dépendent fortement de la nature du signal hôte (image, audio ou vidéo, par exemple). Ainsi, pour un signal hôte vidéo il faut tenir compte d'une contrainte supplémentaire par rapport au signal audio et image. Puisque la dimension spatio-temporelle du signal limite les performances des systèmes de dissimulation. En effet, une marque peut être invisible pour une séquence d'image isolée, mais le défilement de plusieurs séquences différentes marquées au même endroit la rend visible par l'utilisateur.

Invisibilité statistique ou indetectabilité

Certains systèmes de dissimulation de l'information permettent un marquage du contenu sans aucun impacte visuel mais modifient la distribution de probabilité du signal hôte. Ceci peut être problématique surtout dans le cas de stéganographie, puisque l'indetectabilité est une condition. Dans le cas du tatouage numérique robuste, un attaquant pourrait localiser les parties tatouées du signal en procédant à une étude statistique du signal marqué. Il pourra ainsi mener de puissantes attaques ciblées éliminera le tatouage, sans altérer la qualité du signal global.

Certains domaines tels que l'imagerie spatiale ou médicale ne tolèrent aucune distorsion sur le signal hôte. Ainsi, des techniques de data hiding ont été dévellopées pour répondre à ces contraintes. Par exemple, le reversible watermarking est une technique qui permet d'enlever complètement le tatouage inséré et le recouvrement total du signal original (″Reversible watermarking (also known as lossless, distortion-free, invertible) removes completely the watermark and exactly recovers the original signal/image″ [20]). Ainsi, plusieurs travaux ont traité du reversible watermarking pour plus d'information voir les articles suivants [21] [22] [21] [23] [24][20].

Sécurité de la dissmulation de l'information

La Sécurité d'un système de tatouage est définie par la difficulté qu'aura un utilisateur mal-intentionné pour détecter/retrouver l'information insérée ou pour supprimer la marque et récupérer le signal hôte originale. La sécurité d'un système

de dissmulation de l'information repose, en partie, sur une ou plusieurs clefs cryptographiques utilisés à l'insertion et à l'extraction de l'information. Par exemple, dans certains schémas de tatouage, l'information est insérée sous la forme d'un signal produit par un générateur pseudo-aléatoires dont la combinaison d'entrée représente la clef secrète. Il existe deux niveaux de sécurité, dans le premier, un utilisateur non-autorisé ne pourra pas détecter, décoder ou lire le message inséré. Dans le second, le message inséré est crypté et ne peut être lu qu'à l'aide d'une clef secrète. Autrement dit, il faut que le système de dissimulation de l'information soit capable de contrer toute attaque venant d'une personne mal-intentionnée et la clef de cryptage doit rester la dernière défense contre les attaques.
Dans ce travail de thèse, la caractéristique de sécurité prend une importance élevée, puisqu'elle représente la grande différence entre les différentes applications étudiées. De plus, comme il a été relevé par Cayre et al. dans leur article sur la securité du tatouage numérique [25], la sécruité est la propriété la moins étudiée dans un domaine de la dissimulation de l'information.

2.2 Le data hiding comme un problème de communication

La data hiding équivaut à une création d'un canal adjacent au canal de communication pour transmettre une information. Ce qui explique que dans plusieurs travaux [1] [5][26], les schémas de data hiding ont toujours été traités comme des schéma de communication.
Dans le contexte du codage et de la transmission, les termes "Etat du canal" et "Information adjacente" sont équivalents. Puisqu'ils font référence à l'information additionnelle disponible au codeur. En 1958, Shannon [27] proposa un codage de canal optimal utilisant l'information d'états (state information) à l'encodeur d'une manière causal, c'est-à-dire, à un instant t donné l'émetteur ne connais que les séquences émises entre l'instant 1 à l'instant t. Gel'fand et Pinsker [28] ont étudié un canal non-causal où l'émetteur a connaissance de tout le signal avant sa transmission de ce dernier.
Dans la suite de ce document, une catégorie des cannaux de transmission à l'origine d'un certain nombre de techniques de data hiding sera présentée.

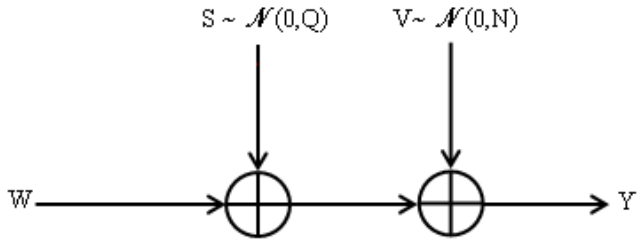


FIGURE 2.2 – Transmission d'un message **w** à travers un canal gaussien.

2.2.1 Canaux avec information adjacente

La modélisation d'un tatouage par un canal gaussien AWGN permet de simplifier l'étude du système. Ainsi, le signal hôte est modélisé par la variable aléatoire S centrée de variance Q, i.e., $S \sim \mathcal{N}(0, Q)$. D'un autre côté, le signal tatouage qui doit être transmit est modélisé par la variable aléatoire W. Puisque le canal de transmission est modélisé par un bruit additif donné par la variable aléatoire V, où : $V \sim \mathcal{N}(0, N)$, alors, le signal reçu peut être formulé comme suit, $Y = S + W + V$ (voir Fig.2.2). La formule de la capacité est donc donnée par l'expression ci-dessous,

$$C = \frac{1}{2} \log_2 \left[1 + \frac{P}{Q + N} \right]. \tag{2.1}$$

Il est clair que, dans ce cas, le signal hôte est considéré comme un bruit ajouté au signal hôte. Cependant, le signal hôte est connu de l'émetteur contrairement au bruit du canal.

2.2.2 Travaux de Costa sur les canaux gaussiens

Présentons d'abord les travaux liés à la communication dans un canal Gaussien avec information adjacente sans aborder la notion de dissimulation d'information. Gel'fand et Pinsker [28] ainsi que Heegard et El Gamal [29] ont montré que la capacité d'un canal sans mémoire avec information disponible à l'encodeur est définie par :

$$C = \max_{Pr(u,w|s)} \{ I(U;Y) - I(U;S) \}, \tag{2.2}$$

avec U est une variable aléatoire représentant le dictionnaire utilisé et $I(.;.)$ est l'information mutuelle entre deux variables aléatoires.
Le cas traité par Costa [14] correspond à la situation vérifiant les conditions suivantes :

1. L'information adjacente ainsi que le bruit du canal sont des variables indépendantes et identiquement distribuées (i.i.d) selon une loi normale.
2. La puissance du signal d'information est bornée par une valeur maximale P, i.e.,
$$E[W^2] \leq P$$

Théorème de Costa

Lorsque la sortie d'un canal gaussien est donnée par : $Y = S + W + V$, où $S \sim \mathcal{N}(0, Q)$ est connu de l'émetteur et $V \sim \mathcal{N}(0, N)$ représente le bruit additif gaussien et le message est donné par le vecteur $\mathbf{w} \in \mathcal{R}^n$, $n \in \mathcal{N}$ satisfait la contrainte sur la puissance $\frac{1}{n}\sum_{i=1}^{n} \mathbf{w}[i]^2 \leq P$. Alors, la capacité C de ce canal est donnée par :

$$C = \frac{1}{2}\log\left[1 + \frac{P}{N}\right] \tag{2.3}$$

Autrement dit, le signal hôte n'a aucune influence sur la capacité du signal.

Preuve On considère $2^{n(I(U;Y)-\epsilon)}$ séquences i.i.d U générées puis distribuées uniformément sur 2^{nR}, où R représente le débit de transmission. Pour chaque séquence U, un index $i(U)$ lui est alloué.
A l'encodage, pour chaque état S et du message W une séquence U est recherchée tel que $(U; S)$ sont des séquences typiquement jointes dans le bloc portant un index : $i(U) = W$. Ensuite, une séquence W est choisie tel que $(W; U; S)$ sont des séquences typiquement jointes qui seront transmises sur le canal. A la réception, le décodeur détermine la séquence unique U qui permet d'avoir $(Y; U)$ typiquement jointes, où Y est la variable aléatoire qui modélise le signal reçu. Enfin, l'index $\hat{W} = i(U)$ est considéré comme une estimation du message W.

Dans son article " Writing on dirty paper " [14], Costa a pris une séquence auxiliaire ayant la forme :

$$U = W + \alpha S, \tag{2.4}$$

où α est un paramètre à déterminer.
L'information sur le dictionnaire U transportée par le signal reçu $Y = S + W + V$

par le récepteur, peut être formulée comme suite :

$$\begin{aligned} I(U;Y) &= H(Y) - H(Y|U) \\ &= H(S+W+V) + H(W+\alpha S) - H(S+W+V; W+\alpha S) \\ &= \frac{1}{2}\log\left(\frac{(P+Q+N)(P+\alpha^2 Q)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\right), \end{aligned} \tag{2.5}$$

De même, l'information sur le dictionnaire de séquence auxiliaires U transportée par la séquence S est donnée par :

$$I(U;S) = \frac{1}{2}\log\left(\frac{P+\alpha^2 Q}{P}\right), \tag{2.6}$$

d'où le débit de la transmission formulé comme suit :

$$R(\alpha) = I(U;Y) - I(U;S) = \frac{1}{2}\log\left(\frac{(P+Q+N)}{PQ(1-\alpha)^2 + N(P+\alpha^2 Q)}\right). \tag{2.7}$$

Cette formalisation permet de gagner en simplicité, puisqu'il suffit de réaliser une maximisation du débit de transmission $R(\alpha)$ par rapport au paramètre α. On obtient :

$$C \triangleq \max_{\alpha} R(\alpha) = \frac{1}{2}\log\left(1 + \frac{P}{N}\right). \tag{2.8}$$

Le paramètre α maximisant le débit est donné par :

$$\alpha = \frac{P}{P+N} \tag{2.9}$$

Cependant et bien que son utilisation a révolutionné le data hiding, le théorème de Costa n'a pas connu un grand succés dans le codage de canal. Ceci vient du fait que dans le codage de canal l'information adjacente représente l'état du canal qui est difficile à évaluer. Par contre, en data hiding le support est lui même l'information adjacente. Nous présentons par la suite, un des premiers systèmes mettant en oeuvre les travaux de costa dans le contexte du data-hiding.

2.2.3 Quantization Index Modulation (QIM)

La quantification est utilisée dans beaucoup de systèmes de tatouage. L'engouement pour cette technique de watermarking vient du fait qu'en présence d'une information adjacente à l'émission, l'encodeur optimal transforme la séquence **s** en

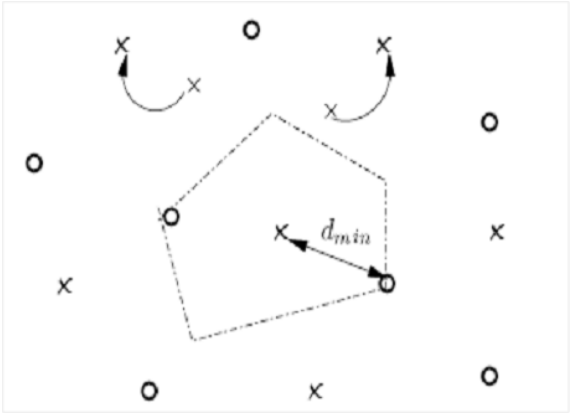


FIGURE 2.3 – Représentation des points de reconstruction des quantificateurs utilisés dans la QIM : Les points de reconstruction représenté par un cercle $''\circ''$ sont ceux du quantificateur modulé par l'index $m = 0$ et les points de reconstruction du quantificateur correspondant à l'index $m = 1$ sont représentés par une croix : $''\times''$.

une séquence auxiliaire $\mathbf{u}$ appartenant au bloc $\mathbf{u_m}$ correspondant au message $\mathbf{m}$. Ce que réalise le quantificateur où pour un quantificateur donné $Q(.)$, chaque point à quantifié $\mathbf{s} \in \mathcal{R}^n$ est déplacé vers le point de reconstruction le plus proche $\hat{\mathbf{s}} = Q(\mathbf{s})$ en additionnant une erreur de quantification $Q(\mathbf{s}) - \mathbf{s}$. Ainsi, pour insérer un message dans un signal anodin, il suffit de choisir un ensemble de quantificateurs puis les indexer à l'aide de l'ensemble des indexes $\mathbf{M}$ à transmettre. Nous obtenons alors $|M|$ quantificateurs , où $|.|$ représente le cardinal d'un ensemble.

Dans l'article ″Quantization index modulation : a class of provably good methods for digital watermarking and information embedding ″ [15], Chen et Wornell ont expliqué les raisons qui font de la quantification l'outil le plus approprié pour le watermarking. De plus ils ont proposé une méthode utilisant *la quantification pour l'insertion de la marque correspond à la modulation par l'index de la Quantification (Quantization index modulation)* plus connu sous le nom : QIM. Elle désigne la technique de tatouage qui module l'index ou la séquence d'indexes avec l'information à transmettre, puis quantifier le signal hôte avec un quantificateur.

Dans le cas de l'insertion d'une information bit par bit ($m \in \{0, 1\}$), chaque échantillon du signal hôte nécessite deux quantificateurs dont les points de recons-

truction sont représentés sur Fig.2.3. L'objectif de la QIM est qu'un échantillon du signal hôte soit déplacé vers l'un des deux types de points de reconstruction, vers $\circ$ si le bit à insérer est $m = 0$, ou alors, vers $\times$ si le bit à insérer est $m = 1$.
Dans le cas d'un canal de transmission sans bruit, les points correspondants aux échantillons hôtes restent à leurs places et le décodeur n'aura aucun mal à reconnaître les points de reconstruction correspondant au bit inséré $m = 0$ ou $m = 1$, puisque la QIM élimine toute les interférences dues au signal hôte.
Dans le cas d'un canal avec bruit et afin d'éviter que les attaques ne déplacent les échatillons du signal hôte, des points de reconstructions d'un quantificateur donné vers les points de reconstructions d'un autre quantificateur, il faut s'assurer que la distance minimale séparant les points d'un quantificateur par rapport à un autre du deuxième, soit supérieure à un certain seuil, afin de resister aux dégradations subit par le signal tatoué pour que le point de reconstruction le plus proche de l'échantillon marqué reste celui correspondant au bit d'information qu'il transporte. Ainsi, pour $|M|$ indexes à transmettre avec, pour cette opération, $|M|$ quantificateurs différents notés par : $\{Q_i(.)\},\ i = 1, ..., |M|$. La distance minimale à respecter pour que le système resiste aux attaques est donnée par [5],

$$d_{\min} \triangleq \min_{(i,j)\in\mathcal{N}:i\neq j} \min_{\mathbf{x}[i],\mathbf{x}[j]} \parallel x[i] - x[j] \parallel, \tag{2.10}$$

où $x[i]$ représente la i^{eme} composante du signal signal tatoué (quantifié avec un quantificateur modulé par le i^{eme} bit d'information).

Dans la situation où les indexes de l'alphabet du message inséré ont une distribution uniforme dans l'ensemble $\mathcal{M}$, le décodeur à distance minimal prend la décision sur le bit index extrait selon la formule suivante :

$$\hat{m} = \min_{m\in M} \parallel \mathbf{y} - Q_m(\mathbf{y}) \parallel \tag{2.11}$$

Le schéma QIM est à la base de tout les systèmes étudiés dans ce rapport.

2.2.4 Distorsion compensation QIM (DC-QIM)

Dans [15] Chen et wornel proposent de diviser le pas de quantification par un facteur $\alpha \in [0, 1[$, pour augmenter la distance minimale qui sépare les points de reconstruction des quantificateurs ceci dans le but de rendre le système plus robuste aux dégradations. Cependant, cette opération multiplie par un facteur $\frac{1}{\alpha^2}$ la distor-

sion due à l'insertion de la marque, ce qui la rend visible. Pour palier à ce problème, la DC-QIM procède à une compensation de cette distorsion en ajoutant une fraction $(1-\alpha)$ de l'erreur de quantification. La séquence marquée pourrait alors être formulée comme suite :

$$\mathbf{x}(\mathbf{s},\mathbf{m}) = Q_{\mathbf{m}}(\alpha\mathbf{s}) - \mathbf{s} \tag{2.12}$$

Bien que le terme ajouté permette de compenser la distorsion due à la division du pas de quantification par le paramètre α, il devient lui même une source d'interférences au récepteur. Pour que le signal de tatouage $\mathbf{w}$ généré à partir du message $\mathbf{m}$ vérifie la contrainte sur la puissance P, la puissance du terme de compensation de la distorsion sera donné par : $(1-\alpha)\frac{P}{\alpha}$. Le rapport watermark sur bruit (w.n.r. :watermark to noise ratio) est alors donné par :

$$w.n.r. = \frac{d_{min}^2/\alpha}{(1-\alpha)^2 P/\alpha^2 + N}. \tag{2.13}$$

La maximisation du w.n.r. par rapport à α permet d'optimiser le système. On obtient alors,

$$\alpha = \frac{P}{P+N}. \tag{2.14}$$

A noter que cette relation coïncide avec le paramètre d'optimisation donné par Costa pour son schéma de communication avec information adjacente.

2.2.5 Scalar Costa Scheme (SCS)

Le SCS [1] est un système basé sur la quantification scalaire et découle directement des travaux de Costa, souvent présenté comme identique au distorsion-compensation QIM [15].
Le schéma de codage proposé par Costa dans son article ≪ Writing on Dirty Paper ≫ est un schéma théorique impossible à réaliser dans la pratique, puisqu'il n'existe pas de dictionnaire ayant un nombre infinis d'éléments. C'est pour cette raison qu'il est appelé le schéma idéal de Costa (ICS : Ideal Costa Scheme). Eggers et al. [1] a proposé de faire une implémentation sous-optimale du Schéma de Costa, dans lequel, il a choisit un dictionnaire qui est le produit de sous dictionnaires. Cette technique est appelée The Scalar Costa scheme (SCS), faisant ainsi référence aux quantificateurs scalaires utilisés lors de la construction des sous-dictionnaires.

L'encodage SCS

Considérons le message $\mathbf{w} \equiv \mathbf{m}$, où $\mathbf{m}$ est la représentation binaire du message (marque) $\mathbf{w}$, est coder avec l'alphabet $\mathbf{d}$ consituée de composantes d tel que : $d \in 0, 1, ..., D-1$, où D est la taille de l'alphabet. La longueur du message est égal à L_x. Généralement, le message est codé en binaire : $D = 2$. Le dictionnaire du schéma de Costa est construit comme le produit de L_x sous-dictionnaires $U^1 : U^{L_x} = \underbrace{U^1 \circ U^1 \circ U^1 ... \circ U^1}_{L_x \ fois}$ .Pour un alphabet constitué de D éléments le sous-dictionnaire est donné par :

$$U^1 = U_0^1 \cup U_1^1 \cup ... \cup U_{D-1}^1. \tag{2.15}$$

Le sous-dictionnaire U^1 dépend du paramètre d'optimisation de Costa α, de la taille D de l'alphabet utilisé pour le codage du message à insérer et du pas de notre quantificateur scalaire à dither. $U^1(\alpha, \Delta, D)$ est équivalent à l'ensemble constitué par les points de reconstructions du quantificateur scalaire uniforme de pas $\frac{\alpha\Delta}{D}$. Autrement dit,

$$U^1(\alpha, \Delta, D) = \left\{ u = l\alpha\Delta + d\frac{\alpha\Delta}{D} \right\}, \tag{2.16}$$

le nombre entier l permet de balayer l'ensemble des points de reconstruction du quantificateur scalaire de pas $\frac{\alpha\Delta}{D}$, alors que d engendre un décalage du quantificateur. Pour utiliser le schema de costa, il faut chercher une paire $(\mathbf{u}, \mathbf{s})$ typiquement jointe. Ce qui est équivalent à rechercher une séquence typique $\mathbf{q} = \frac{\mathbf{w}}{\alpha} = \frac{\mathbf{u}}{\alpha} - \mathbf{s}$ quasiment orthogonal à $\mathbf{s}$. Rechercher $\frac{\mathbf{u}}{\alpha}$ revient à rechercher la quantification de s. Il est donc possible de résumer l'opération de la génération du tatouage dans la formule suivante :

$$\mathbf{q} = Q_\Delta \left\{ \mathbf{x} - \Delta\frac{\mathbf{d}}{D} \right\} - \left\{ \mathbf{x} - \Delta\frac{\mathbf{d}}{D} \right\}, \tag{2.17}$$

où $Q_\Delta(.)$ représente le quantificateur scalaire uniforme de pas Δ.
La séquence d'information à transmettre est formulée par :

$$\mathbf{w} = \mathbf{u} - \mathbf{s}, \tag{2.18}$$

d'où la séquence du signal marquée :

$$\mathbf{x} = \mathbf{s} + \mathbf{w} = \mathbf{s} + \alpha\mathbf{q}. \tag{2.19}$$

Le tatouage SCS dépend de deux paramètres : le pas de quantification Δ et le

facteur d'échelle α. Pour une puissance du watermark donnée σ_W^2, on a :

$$\alpha = \sqrt{\frac{\sigma_W^2}{\Delta^2}} = \frac{\sigma_W^2\sqrt{12}}{\Delta} \tag{2.20}$$

Le décodeur SCS

Il existe plusieurs similitudes entre l'encodage et le décodage dans le système SCS, puisque comme pour l'encodeur, le signal d'entrée du décodeur : $\mathbf{y} = \mathbf{s} + \mathbf{w} + \mathbf{v}$ est quantifié. Cette opération sert à chercher le bloc U correspondant à la quantification du signal reçu $\mathbf{y}$, afin de déterminer l'information qu'il transporte. On procède dans ce cas à un décodage à décision dure.
Soit $\mathbf{r}$ le critère de décision qui permet de déterminer le bloc correspondant au signal reçu $\mathbf{y}$. Le crière $\mathbf{r}$ est donné comme suite :

$$\mathbf{r} = Q_\Delta(\mathbf{y}) - \mathbf{y} \tag{2.21}$$

Pour un système SCS binaire, le i^{eme} bit d'information $\mathbf{m}[i]$ est déterminé comme suite :

$$\mathbf{m}[i] = \begin{cases} 1 & : \mid \mathbf{r}[i] \mid < \frac{\Delta}{2} \\ 1 & : \mid \mathbf{r}[i] \mid \approx \pm\frac{\Delta}{2} \end{cases} \tag{2.22}$$

2.2.6 Trellis coded quantization (TCQ)

Le schéma TCQ peut être considéré comme une variante (amélioration) du SCS. Il utilise plusieurs sous-dictionnaires obtenu grâce au partitionnement pseudo-aléatoire de l'espace. Ceci est rendu possible grâce à l'utilisation d'un trellis.
Soit le treillis de Fig.2.4. Pour l'utiliser en dissimulation de l'information, on considére que les transitions d'un état à un autre, représentées par un trait plein pour un bit message 1 et celle en trait discontinus correspondent à un bit message égal à 0. Un sous dictionnaire est alloué à chaque transition. Pour insérer un bit message, il suffit de choisir la transition correspondante au bit d'information et de choisir un mot de code du sous-dictionnaire correspondant, ceci en tenant compte de l'échantillon hôte. Pour chaque message à insérer correspond un chemin du trellis et un dictionnaire composé des sous-dicionnaires de chaque transition du chemin choisit. Ainsi, le chemin du trellis de Fig.2.4 représenté avec un trait vert correspond au message binaire : 010 et au dicionnaire : $C = C_0, C_2, C_0$.

La tatouage TCQ est une quantification qui utilise un treillis associé à un dictionnaire structuré, elle permet de réduire la complexité du système de tatouage tout en diminuant la distorsion. Cette approche pour le tatouage est considérée comme une variante de celle utilisée en codage (codage de canal puis codage de source [30]). La TCQ combine donc un ensemble de treillis avec un partitionnement des ensembles de la TCM (Treillis Coded Modulation) [31] pour diminuer la distorsion et la complexité du système.

Quantification codée en treillis en codage de canal

La figure Fig.2.4 représente un exemple de treillis à 4 états, où chaque branche est associée à un sous-dictionnaire . Pour construire le dictionnaire de la TCQ, il est défini un ensemble C constitué des points de reconstructions d'un quantificateur scalaire est de taille $2^{R+R'}$ (R représente le taux de codage et R' désigne le nombre de bits qui spécifient les mots de codes choisit dans le sous-dictionnaire). Dans le système de Fig.2.4, R est égale à 2 bits par échantillons (bit per sample, bps), alors que R' est égale à 1. Pour un codage à $2bits/seconde$, C est deux fois plus large que le quantificateur scalaire correspondant. Après sa construction C est été construit, il est divisé en sous-ensembles contenant chacun des mots de code. Lors de la quantification, le mot de code le plus proche de l'échantillon du signal à coder, est déterminé pour chaque sous-ensemble. Afin de rendre possible l'attribution de valeurs à chaque brache, Viterbi [32] est utilisé par la suite pour déterminer le chemin du treillis minimisant la distorsion, on obtient ainsi la suite des mots à utiliser ainsi que les différentes transitions à effectuer dans le treillis.

Quantification codée en treillis pour le watermarking

Dans le cas du tatouage numérique, les chemins de treillis son forcés par les valeurs des bits du message $\mathbf{m}$. Les échantillons du signal hôte $\mathbf{s}$ sont quantifiés à l'aide des dictionnaires correspondants au chemin emprunté, ce qui donne un débit de 1 bit par échantillon.
Pour utiliser la TCQ en data hiding, ce sont les transitions du treillis qui déterminent les bits d'informations insérés (voir la figure Fig.2.4). Là aussi, un sous-dictionnaire est alloué à chaque transition. Pour insérer un bit message, il suffit de choisir la transition correspondant au bit d'information et de choisir un mot de code du sous-dictionnaire correspondant, ceci en tenant compte de l'échantillon hôte. Chaque

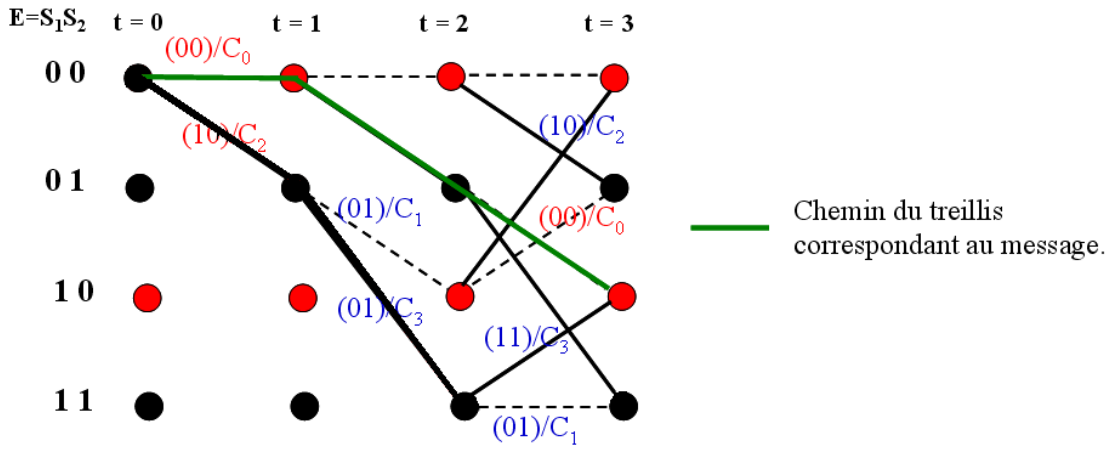


FIGURE 2.4 – Insertion d'un message binaire $\mathbf{m} = [010]$ dans signal hôte à l'aide du treillis. Les transitions représentées par un trait plein correspondent à un bit message 1 et celle en trait discontinus correspondent à un bit message égal à 0.

message à insérer correspond à un chemin du trellis et un dictionnaire composé des sous-dicionnaires de chaque transition du chemin choisit. Ainsi, le chemin du trellis de Fig.2.4 représenté avec un trait vert correspond au message binaire : 010 et au dicionnaire : $C = C_0, C_2, C_0$.

Au décodage, l'algorithme de Viterbi est utilisé pour retrouver le meilleur chemin parmi tous les chemins du treillis possibles, les transitions de ce chemin permettront de récupérer le message inséré.

Il existe une autre approche pour l'utilisation de la TCQ dans le watermarking appelée : TCQ Initial State (TCQ-IS) [33], dans laquelle le message est inséré dans l'état initial du chemin du treillis des transitions. Bien que la TCQ-IS sont plus robuste que la TCQ classique, son utilisation n'est justifiée que pour les deux raisons suivantes :

- La taille du message dans le cas d'un tatouage avec la TCQ-IS, dépend uniquement du nombre d'états du treillis utilisé pour la quantification TCQ. L'insertion d'un message de taille importante revient donc à choisir un treillis ayant un très grand nombre d'états. Ceci entraine une augmentation de la complexité du système de tatouage.
- Le gain en termes de robustesse du système TCQ-IS par rapport au TCQ dépend du canal de transmission ce gain généralement peu important.

2.3 Spread Transform (ST)

Chen et Wornel [15] ont introduit une apporche génerale pour le tatouage numérique robuste. Ils proposaient d'étaler l'infomation à insérer sur plusieurs échantillons, en procédant à une transformation/transformation inverse du signal hôte. La transformation du signal hôte habituellement utilisée, telle que décrite la dans la bibiliographie [1][15], est donnée par :

$$\mathbf{s}^{\mathrm{st}}[l] = \sum_{i=\tau l}^{\tau l+\tau-1} \mathbf{s}[i] \times \mathbf{t}[i], \text{ with } \tau \in \mathbb{N}^{\star}. \tag{2.23}$$

tel que les $\mathbf{s}[i]$ sont les échantillons non-transformés et les $\mathbf{t}[i]$ représentent les paramètres d'étalement.
Cependant, le terme *transformation* peut être trompeur, puisque l'opération décrite par Eqn.2.23 est plus proche d'une projection suivant une direction $\mathbf{t}$ qu'une transformation. Le signal composite $\mathbf{x}$ est donné comme suit :

$$x[i] = \mathbf{s}[i] + (\mathbf{u}_m^{ST}[l] + \mathbf{s}^{st}[l]) \cdot \mathbf{t}[i], \tag{2.24}$$

où $\mathbf{u}_m^{ST}[l]$ est le l^{me} mot de code du dictionnaire, correspondant au message, du signal hôte $\mathbf{s}^{st}[l]$.
Le ST a de nombreuse caractéristiques interessantes surtout en termes de robustesse et d'imperceptibilité. Les auteurs dans l'article [1] proposent la formule suivante,

$$wnr = wnr_{\tau} - 10\log_{10}(\tau), \tag{2.25}$$

où wnr correspond à Watermark to Noise Ratio, ou, le rapport puissance du tatouage à bruit, alors que wnr_{τ} représente le wnr mais dans le domaine transformé. Dans ce rapport, une nouvelle approche concernant le ST sera discutée utltérieurement. Le principe de fonction du ST est résumé dans Fig.3.8.

2.3.1 Performances

Nous avons pris comme réference le systèmes SCS [1], dans ce travail de thèse, parce que ce système est un schéma basique des systèmes basés sur la quantification et il est très perfomant (voir [1] ou [15]), surtout, en termes de capacité et de robustesses, par rapport, aux systèmes non-informés, en particulier, le fameux spread

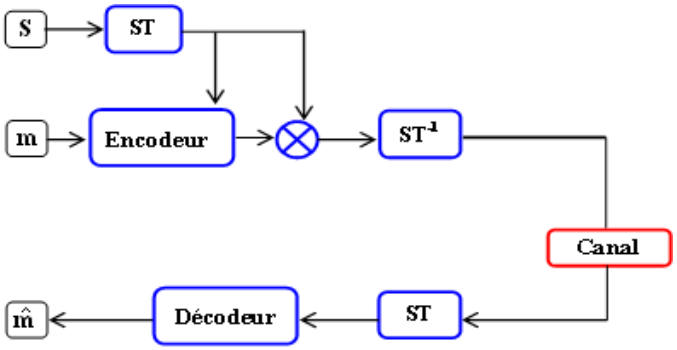


FIGURE 2.5 – Spread transform combiné avec des systèmes de dissimulation d'information informé.

spectrum watermarking [26].

Fig.2.6 montre l'avantage des systèmes (SCS et ICS) par rapport au très connu système de tatouage SS, ce ci est dû au fait de considérer le signal hôte comme information adjacente utilisée pour transmettre l'information et non pas comme un bruit additif qui dégrade l'information dissimulée avant même d'être envoyée. On note un avantage du ICS par rapport au SCS, il dû au fait que le ICS est un système idéale qui donne de très bon résultats théoriques mais qui ne peut être utilisé à cause de l'hypothèse d'un dictionnaire infini. D'un cuatre côté, Fig.2.7 justifie en partie notre choix du ST comme système d'optimisation des systèmes de dissmulation del'information, puisqu'elle montre l'avantage d'uiliser le Spread Transforme (ST) par rapport au répetition coding, de plus, il démontre que la robustesse apportée par le ST n'est par dûe uniquement au fait de répeter l'information plusieur fois mais bien grâce à l'étalement particulier du ST qui permet une meilleur amplification de l'information à l'entrée du décodeur.

fingerprinting.

Durant cette thèse, nous nous sommes concentré sur les principales application des systèmes de data-hiding, en l'occurence, la stéganographie dans un contexte warden actif, un cas plus générale que la stéganogranohie classique tel qu'il est décrit dans [5], le tatouage robuste et à un degré moindre le

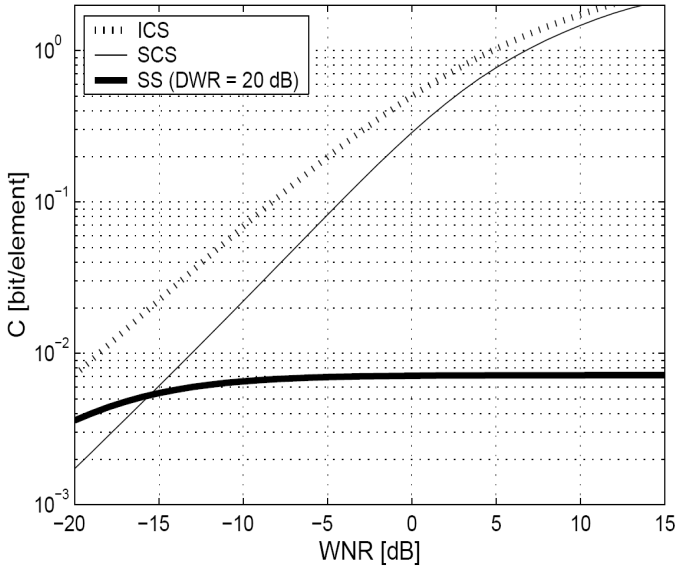


FIGURE 2.6 – Capacité d'un schéma non informé : Spread Spectrum watermarking et de schémas informés basés sur les travaux de Costa : Scalar Costa Scheme (SCS) et Ideal Costa Scheme (ICS)) [1].

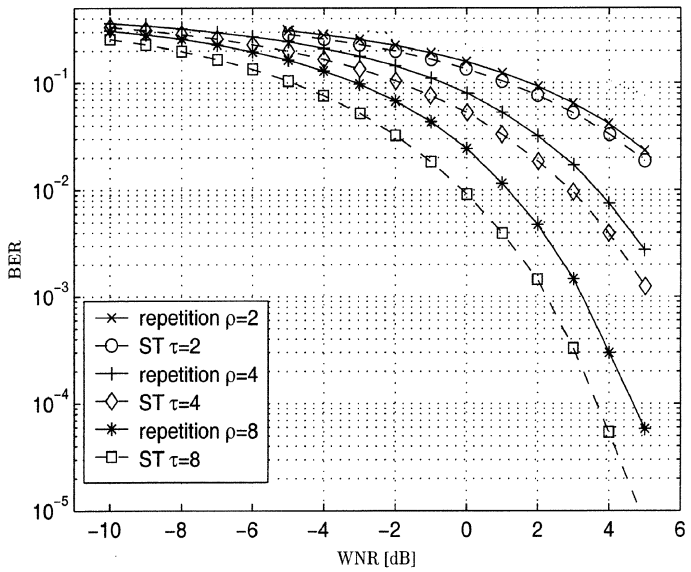


FIGURE 2.7 – b.e.r. du SCS avec le le codage à répetition et le tatouage ST-SCS pour des facteur d'étalement identique ($\rho = \tau$) [1].

Chapitre 3

Stéganographie

Ce chapitre présente une partie de nos travaux effectués dans le domaine de la stéganographie. Le problème traité est l'utilisation de systèmes de tatouages informés basés sur la quantification en stéganographie. Le contexte de notre travail est la stéganographie avec gardien actif qui n'est autre que le cas le plus courant, stéganographie gardien passif, mais avec des contraintes en plus que nous détaillerons dans ce chapitre. Notez qu'avant que ce travail soit effectué, les systèmes informés, en particulier ceux basés sur la quantification, souvent été considérés comme incompatible dans un contexte stéganographique.

3.1 Introduction

L'objectif principal de la stéganographie est l'indétectabilité. Ceci signifie que le gardien, appelé Wendy, dans le problème des prisonniers de Simmon [34] ne peut décider s'il existe un message -envoyé par Alice à Bob- est présent ou pas dans le stégo-document. Ce problème d'évaluation de l'habilité du gardien à détecter la présence du message caché ou pas est formalisé, dans la suite de ce chapitre, en se basant sur les statistiques du stégo-signal. Dans ce chapitre, nous utilisons l'entropie relative 1-Dimensions et 2-Dimensions comme métrique pour l'indétectabilité des stégo-systèmes. Le contexte de ce travail est la stéganographie avec gardien actif, i.e. le gardien a la possibilité d'agir pour empêcher le Bob de recevoir le message caché. Alice et Bob utilisent un certain paramètre secret pour communiquer et Wandy va tenter d'agir mais d'une manière aveugle, sans aucune connaissance à priori sur les paramètres du stégo-système. Nous modélisons pour ce travail toutes les attaques

aveugles du gardien par un bruit blanc Gaussien additif AWGN (Additive White Gaussian Noise). Le choix de ce modèle est dû au grand nombre d'attaques possibles sur le stégo-signal qui sont souvent combinées pour être efficaces, en supposant que les attaques respectent les conditions du theorème Central Limit [35], le resultat de la somme de ces attaques converge vers un ajout de bruit AWGN. Ainsi, les stégo-systèmes sont considérés comme des canaux de communication contraints par l'indétectabilité et les attaques du gardien.
Dans cette partie, nous procéderons à une étude stéganographique des stégo-systèmes informés : Schéma Scalaire de Costa ou SCS (Scalar Costa scheme), Schéma scalaire de Costa avec clef secrète, quantification codé par treillis ou TCQ (Trellis Coded Quantization) et l'étalement transformé du schéma scalaire de Costa ou ST-SCS (Spread Transform Scalar Costa Scheme). Des formulations théoriques, ainsi, que des évaluations expérimentales démontrent les avantages et les limites de chaque schéma en termes d'indétectabilité et de capacité. De plus, nous proposons un nouveau stégo-schéma basé sur la combinaison du ST et du TCQ. Nous montrons que le schéma proposé permet une meilleure résistance face aux attaques du gardien actif Wendy et permet un bon compromis entre l'indétectabilité, la capacité et la résistance face aux attaques du gardien Wendy.

3.2 Principes de base

En se basant sur les définitions données dans le Livre de Cox et al.[5], le contexte du gardien actif veut dire,

- Wendy analyse le stego-signal dans le but de détecter un éventuel stégo-message (stéganographie avec gardien passif),
- le gardien Wendy altère tout les contenus dans l'espoir de détruire un éventuel stego-message,
- le gardien n'est pas malicieux (pour plus de détails sur la définition du gardien malicieux voir [5]). Les actions du gardien ne sont pas basées sur les spécificités du stego-schéma. Par exemple, Wendy ne pourra pas procéder à ce qui appelé "3-Delta attack" , décrite dans l'article [36], dans le cas des stégo-systèmes basés sur la quantification, puisque des valeurs particulières du pas de quantification doivent être utilisées pour réussir l'attaque.

Dans ce qui suit, nous supposons qu'Alice et Bob utilisent une stégo-clef secrète pour communiquer et Wendy n'a pas de connaissance sur cette clef, ni sur les paramètres secrets partagés entre les deux détenus. Afin de proposer une analyse détaillée des schémas qui peuvent être utilisé dans le contexte de ce chapitre, procédons à la définition des propriétés de bases des systèmes de stéganographie :

3.2.1 Indétectabilité

C'est le but principal de la stéganographie, Cox et al. [5] définissent l'indétectabilité comme *l'impossibilité de détecter la présence du stego-message inséré dans un document*, i.e. le gardien ne peut faire la différence entre un stego-document et un document innocents. Dans l'article [37], la contrainte d'indétectabilité parfaite est donnée par : $p_S = p_X$, où p_S est la p.d.f. du cover-signal $\mathbf{s}$ et p_X est la p.d.f. du stego-signal $\mathbf{x}$. Puisqu'il est difficile d'obtenir l'indétectabilité parfaite et dans le but de comparer les performances des stego-systèmes en termes d'indétectabilité, nous mesurons le niveau de cette dernière à l'aide de l'entropie relative $D(p_S||p_X)$. Ainsi, notre objectif serait de minimiser cette distance (bien que ce n'est pas vraiment une distance !), i.e., le stego-signal $\mathbf{x}$ devrait respecter la condition suivante :

$$D(p_S||p_X) = \int_{-\infty}^{+\infty} p_S(z) \ln \frac{p_S(z)}{p_X(z)} \, \mathrm{d}z \leq \epsilon, \tag{3.1}$$

où ln(.) est le logarithme népérien et ϵ est une valeur réelle positive très petite.

3.2.2 Transparence (fidélité)

Dans l'article [37], les auteurs définissent la transparence (aussi appelée l'imperceptibilité) comme *le taux de rapprochement du cover-text du stego-texte sous une métrique de distorsion (fidélité) appropriée*. la metric considérée dans ce travail est la distorsion d'insertion avec une puissance du signal hôte fixée et le Peak Signal-to-Noise ratio (PSNR) est utilisé pour l'évaluation de cette fidélité lors des expériences sur les images réelles. Notons que les critères subjectifs ne sont pas utilisés dans ce travail, puisqu'ils utilisent souvent le modèles visuel humain HVS (Human Visual System [38]), -pour les images et vidéos-, ou les modèles du système auditif humain HAS (Human Auditory System) [39] pour les signaux audio.

Généralement et contrairement au tatouage robuste, la fidélité n'est pas consi-

dérée comme importante pour la stéganographie. Alice est libre de choisir le cover-signal, puisque celui-ci n'a aucune valeur dans le contexte de la stéganographie. Cependant, est-il correct de dire qu'il n'existe aucune corrélation entre la fidélité et l'indétectabilité ?

Notons que dans le cas spécifique de la stéganographie sur les images, Eggers et al. [1] a écrit : *La relation entre la distorsion d'insertion et la sécurité des schémas stéganographique n'est pas incluse dans les travaux de Cachin, cependant, elle est cruciale pour la stéganographie appliquée aux images.* Dans notre contexte (qui inclut la stéganographie appliquée aux images), nous allons déterminer la relation entre l'entropie relative $D(p_S||p_X)$ -entre le cover-signal $\mathbf{s}$ considéré comme gaussien et le stego-signal $\mathbf{x}$- et la puissance d'insertion. Dans [40] les auteurs ont procédé un développement intéressant de la stégo-sécurité au sens de Cachin pour le stego-system SCS (avec et sans la clef secrète). Dans ce travail, nous évaluons théoriquement l'entropie relative entre le cover et le stégo-signal qui pourrait être aussi une généralisation de l'évaluation de la stégo-sécurité selon Cachin pour les stégo-systèmes basés sur la quantification et qui sont étudiés dans ce chapitre. Notre objectif par nos développements est l'évaluation théorique de la stégo-sécurité pour des signaux gaussiens et pour donner une sorte d'estimation de la relation entre la stégo-sécurité et la puissance d'insertion.

Ce travail concerne les systèmes de dissimulation des données avec information adjacente qui sont basés sur les travaux de Costa [14]. Afin de réccupérer les résultats théoriques dévellopés dans ces travaux, nous respecterons les conditions établies dans l'article [14]. Ainsi, nous supposons dans la suite de ce document (sauf indication contraire) que le cover-signal $\mathbf{s}$ suis une loi Gaussienne de variance σ_S^2. Aussi, nous assumons que le stégo-système utilisé préserve la "Gaussiennetée" du stégo-signal. Ce dernier a une variance $\sigma_X^2 = \sigma_S^2 + \sigma_E^2$, où σ_E^2 est la variance du signal d'insertion E (pas nécessairement Gaussien). Les hypothèses établies sont plus réalistes que cela pourrait sembler. Par exemple, si nous prenons le schéma SCS avec clef secrète [1] (parfois appelée dithering) et dans le cas où le quantificateur scalaire vérifie les conditions de hautes résolutions [41], toutes les hypothèses seront vérifiées. Ainsi, grâce à la haute résolution du quantificateur scalaire, l'erreur de quantification, utilisée pour la construction du signal d'insertion, est indépendante du cover-signal (information adjacente) (pour plus de détails voir [41]). Par ailleurs, Eggers et al. dans leur article [42] ont traité cette question dans le cas du SCS, où ils ont écrit : "*...Le signal watermark du SCS a une distribution uniforme de largeur* $\alpha\Delta$ *et est*

statistiquement indépendant du cover-signal x''. De plus, la clef secrète (ou dithering) permet de limiter les distorsions sur la p.d.f. du stégo-signal, tel qu'il a été montré dans [2], ce qui laisse le stégo-signal "Approximativement" Gaussien (lorsque le cover-signal est Gaussien). Sous les hypothèses précédentes, l'entropie relative est donnée par la formulation suivante :

$$D_{theo}(p_S||p_X) = -\frac{1}{2}\left[\frac{1}{1+\sigma_S^2/\sigma_E^2} - \ln\left(1+\sigma_E^2/\sigma_S^2\right)\right], \tag{3.2}$$

tel que,

$$\frac{\partial D_{theo}(p_S||p_X)}{\partial \sigma_E^2} = \frac{1}{2}\frac{\sigma_E^2}{(\sigma_E^2+\sigma_S^2)^2} > 0. \tag{3.3}$$

Preuve *Afin de réccupérer les résultats sur les schémas de dissimulation des données informés, on gardera dans ce dévellopement les mêmes conditions établies dans l'article d'Eggers et al. [1]. Ainsi, le stégo-signal est considéré comme un ensemble de réalisations de variables aléatoires Gaussiennes, indépendantes et non stationnaires :* $\mathcal{X} = \{X[1], \ldots, X[N]\}$. *Considérons la formulation suivante :*

$$X[i] = S[i] + E[i], \ \forall i = 1, \ldots, N, \tag{3.4}$$

où S *représente le cover-signal modélisé par l'ensemble de réalisations des variables aléatoires Gaussiennes, indépendantes et non stationnaires :* $\mathcal{S} = \{S[1], \ldots, S[N]\}$. *Le signal inséré* E *est modélisé par l'ensemble des réalisations de variables aléatoires, indépendantes et non stationnaires :* $\mathcal{E} = \{E[1], \ldots, E[N]\}$. *Notons que le signal d'insertion* E *n'est pas nécessairement Gaussien, ainsi, le développement qui suit n'est pas celui d'une entropie relative entre deux signaux Gaussien, même s'il y avait des points communs. L'entropie relative entre les p.d.f. du cover-signal et du stégo-signal est donnée comme suit :*

$$D(p_S||p_X) = \int p_S(z) \ln \frac{p_S(z)}{p_X(z)} \, dz. \tag{3.5}$$

Si nous considérons que le cover-signal $S \sim \mathcal{N}(0, \sigma_S^2)$ *et le stégo-signal restent Gaussiens, centrés et de variance égal à* $\sigma_S^2 + \sigma_E^2$, *où* σ_E^2 *représente la variance du signal d'insertion (le signal d'insertion est considéré comme indépendant du cover-signal* S*), donc, l'entropie relative théorique entre les p.d.f. du cover et stégo-signal est*

donnée par la formule suivante :

$$\begin{aligned} D_{theo}(p_S||p_X) &= \int_{-\infty}^{+\infty} p_S(z) \ln \left(\frac{\frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-\frac{-z^2}{2\sigma_S^2}}}{\frac{1}{\sqrt{2\pi\left(\sigma_S^2+\sigma_E^2\right)}} e^{-\frac{z^2}{2\left(\sigma_S^2+\sigma_E^2\right)}}} \right) \, dz \\ &= \int_{-\infty}^{+\infty} p_S(z) \ln \left(e^{-\frac{1}{2}\left(\frac{z^2}{\sigma_S^2}-\frac{z^2}{\sigma_S^2+\sigma_E^2}\right)} \sqrt{\frac{\sigma_S^2+\sigma_E^2}{\sigma_S^2}} \right) \, dz \\ &= \int_{-\infty}^{+\infty} p_S(z) \ln \left(\sqrt{\frac{\sigma_S^2+\sigma_E^2}{\sigma_S^2}} \right) \, dz - \int_{-\infty}^{+\infty} p_S(z) \frac{z^2}{2} \left(\frac{\sigma_E^2/\sigma_S^2}{\sigma_S^2+\sigma_E^2} \right) \, dz \\ &= \ln \left(\sqrt{\frac{\sigma_S^2+\sigma_E^2}{\sigma_S^2}} \right) \int_{-\infty}^{+\infty} p_S(z) dz - \left(\frac{\sigma_E^2/\sigma_S^2}{\sigma_S^2+\sigma_E^2} \right) \int_{-\infty}^{+\infty} p_S(z) \frac{z^2}{2} dz \end{aligned} \tag{3.6}$$

Puisque $\int_{-\infty}^{+\infty} p_S(z) \, dz = 1$ *et* $\sigma_S^2 = \int_{-\infty}^{+\infty} z^2 p_S(z) \, dz$*, donc,*

$$D_{theo}(p_S||p_X) = \frac{1}{2} \ln \left(\frac{\sigma_S^2+\sigma_E^2}{\sigma_S^2} \right) - \frac{1}{2} \frac{\sigma_E^2}{\sigma_S^2+\sigma_E^2}. \tag{3.7}$$

D'après Eqn.3.2 et Eqn.3.3 l'entropie relative $D_{theo}(p_S||p_X)$ est une fonction strictement croissante par rapport à la puissance d'insertion. Puisque cette dernière affecte directement la fidélité (l'imperceptibilité), l'indétectabilité et la puissance d'insertion ont théoriquement le même comportement (sorte de proportionnalité) : si la puissance d'insertion diminue l'indétectabilité devient meilleure, dans le cas contraire, elle se dégrade. Ceci est au moins vrai pour les stégo-systèmes étudiés dans ce chapitre. Si D_stat représente le maximum de puissance d'insertion qui permet l'indétectabilité offerte par Eqn.3.1, le signal d'insertion qui vérifierait les hypothèses précédentes avec une puissance d'insertion infèrieur à D_{stat} sera indétectable. Prenons la métrique suivante :

$$D_1 = \min(D_{\text{stat}}, D_{\text{fid}}), \tag{3.8}$$

où D_{fid} est la puissance d'insertion maximale qui assure l'imperceptibilité. Ainsi, si la puissance d'insertion σ_E^2 vérifie la condition

$$\sigma_E^2 \leq D_1, \tag{3.9}$$

alors, le processus d'insertion respecte les contraintes d'indétectabilité et de fidélité.

3.2.3 Capacité

Dans le contexte de la stéganographie avec un gardien passif, où la seule contrainte est celle de l'indétectabilité. Cox et al. [5] définissent la capacité comme "*le nombre maximum de bits d'information que l'on pourrait cacher dans un cover-document, tel que la probabilité de détection par l'adversaire est négligeable.*". Tel qu'il a été décrit dans la sous-section précedente la contrainte d'indétectabilité est vérifiée si $\sigma_E^2 \leq D_{stat}$. Par ailleurs et afin d'améliorer les performances des stégo-systèmes, la condition donnée par Eqn.3.9 qui est plus contraignante sera adoptée. Ainsi, nous pourrons respecter les deux conditions : fidélité et indétectabilité. Cependant, le contexte de ce chapitre est la stéganographie avec gardien actif. Ceci modifie la définition de la capacité stéganographique donnée dans le contexte du gardien passif. En d'autres termes, le gardien définie le canal de transmission. Le gardien actif signifie :

- la communication entre Alice et Bob doit rester indétectable : la seule contrainte du gardien passif,
- Wendy, le gardien, insert des modifications avec une puissance maximale égale à D_2 : toutes les modifications possibles sont modélisées par une attaque AWGN avec une puissance égale à σ_V^2,
- Wendy n'utilise pas les particularités et les spécificités du stégo-signal pour effectuer ses attaques. Par exemple, elle ne va pas utiliser des attaques qui utilisent le pas de quantification dans les schémas d'insertion basés sur la quantification (SCS, TCQ, ... etc). Autrement, le gardien deviendra "malicieux" ce qui sera un contexte différent du gardien actif adopté dans cette partie du travail tel qu'il est montré dans l'article [5].

Le schéma général du stégo-système adopté dans cette partie du travail est décrit sur la figure Fig. 3.1. Le cover-signal est modélisé par l'ensemble des séquences : $\mathcal{S} = \{S_1, \ldots, S_N\}$ des échantillons i.i.d. décrient à partir de la p.d.f. $\{p_S(s), s \in \mathcal{S}\}$. Le message $\mathbf{m} \in \mathcal{M}$ est celui qui est inséré dans le signal $\mathbf{s}$. L'encodeur produit le stégo-signal $\mathbf{x}$, dans le but de transmettre le message $\mathbf{m}$ au décodeur. Ainsi, le gardien observe le signal $\mathbf{x}$ et teste si ce signal suit la p.d.f. du cover-signal p_S. Si ce n'est pas le cas, le gardien met fin à la communication entre Alice et Bob. Lorsque le stégo-signal $\mathbf{x}$ ne présente pas de suspicion, le gardien procède à une distorsion de signal et produit le signal corrompu $\mathbf{y}$ en faisant passer $\mathbf{x}$ à travers un certain canal

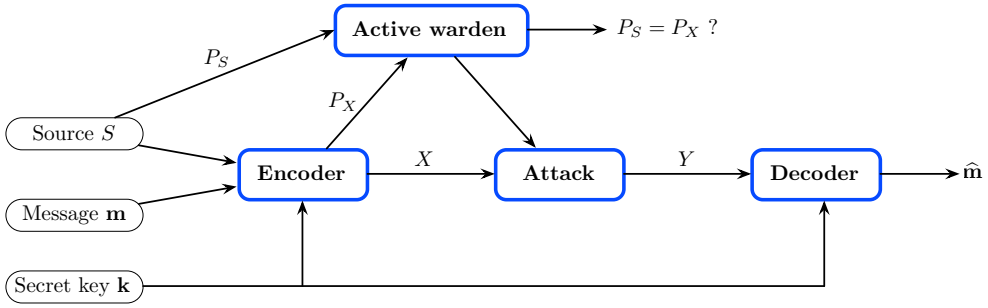


FIGURE 3.1 – Schéma de la stéganographie dans un contexte de gardien actif comme un schéma de communication.

d'attaque : $p_{Y|X}(y|x)$. En d'autres termes, lorsque la contrainte d'indétectabilité est vérifiée, le stégo-signal sera forcément distordu.

Stégo-capacité Dans ce qui suit, nous développons une formule de la stégo-capacité en se basant sur les hypothèses établies précédemment et les contraintes imposées par le contexte de ce travail. Pour cette raison, nous utilisons les définitions de la capacité d'un canal de communication et la stégo-capacité.
Supposons que la p.d.f. du signal de sortie dépend uniquement du signal d'entrée à un instant donné (système de communication sans mémoire). Ainsi, il est possible de considérer le gardien actif comme un canal discret. D'après l'article [43], la capacité d'un canal de communication est donnée par la formulation suivante :

$$C = \max_{p_X(x)} I(X;Y), \tag{3.10}$$

où le maximum est pris sur toute les distributions possibles de l'entrée $p_X(x)$.
En stéganographie, le canal est défini par le gardien Wendy. Ceci modifie la définition même de la capacité de canal classique. D'après l'article [37], la capacité stéganographqiue $C^{\text{stego}}(D_1, D_2)$ comme le supremum de tous les taux atteignables, i.e., le taux R est atteignable si $|\mathcal{M}| \geq 2^{NR}$ et $\sup_{p_{Y|X}} p_e \to 0$ tel que $N \to \infty$, où p_e est la probabilité d'erreur. Donc, la capacité stéganographique est donnée par :

$$C^{\text{stego}} = \lim_{L\to\infty} \max_{p_{XU|S}\in\mathcal{Q}(L,p_S,D_1)} \min_{p_{Y|X}\in\mathcal{A}(p_X,D_2)} I(U;Y) - I(U;S), \tag{3.11}$$

où :

– L est un entier arbitraire très grand qui définit la taille de l'alphabet $\mathcal{U} =$

$1, 2, \ldots, L$ pour une variable auxiliaire aléatoire U dans le stégo-système avec une information adjacente.

– $P_{XU|S}(x, u|s)$ est le canal stéganograhique subissant une distorsion D_1 et une fonction de distorsion $d(s, x)$ (voir [37] pour une définition plus détaillée) dont la marginale conditionnelle $P_{X|S}$ appartient à l'ensemble

$$\begin{aligned}\mathcal{Q}^{\text{stego}}(p_S, D_1) &= \{p_{X|S} : \sum_{s,x} p_{X|S}(x|s)p_S(s)d(s, x) \leq D_1, \\ &\quad p_X(x) = \sum_s p_{X|S}(x|s)p_S(s) = p_S(x), \forall x \in \mathcal{S}\}, \quad (3.12)\end{aligned}$$

donc, $\mathcal{Q}^{\text{stego}}(p_S, D_1)$ est l'ensemble des canaux soumis à une distorsion D_1. Notons que les éléments de cet ensemble défini dans Eqn.3.12 vont être restreints à ceux correspondant aux stégo-systèmes informés.

– Nous noterons par :

$$\mathcal{A}(p_X, D_2) = \left\{ p_{Y|X} : \sum_{x,y} p_{Y|X}(y|x)p_X(x)d(x, y) \leq D_2 \right\},$$

l'ensemble de toutes les attaques faisables sur un canal discret sans mémoire.

D'un autre côté, le contexte de cette partie du travail est celui du gardien actif. Toutes les attaques de Wendy sont modélisées par une attaque AWGN, tel que la puissance du bruit est égale à $D_2 = \sigma_V^2$, où σ_V^2 est la variance du bruit additif. Par ailleurs, nous fixons la p.d.f. conditionnelle $p_{Y|X}$ et sa maximisation, sur tous les canaux discrets sans mémoire dans Eqn.3.11, n'est pas nécessaire. De plus, nous supposons que le stégo-signal vérifie la condition donnée par Eqn.3.9. Donc, la stégo-capacité donnée par Eqn.3.11 converge vers la définition de la capacité donnée par Gel'fand et Pinsker [44] et devient alors :

$$C = \max_{p_{XU|S}} \{I(U; Y) - I(U; S)\}, \quad (3.13)$$

où U est une variable auxiliaire. Les schémas qui nous intéressent dans ce travail sont tous basés sur les travaux de Costa et, comme dans l'article [1]. D'après Eqn.3.11, Eqn.3.12 et Eqn.3.10 la stégo-capacité dans le contexte de chapitre devient :

$$C^{\text{stego}} = C = \max_{\alpha} I(Y; M), \quad (3.14)$$

où α est le paramêtre de Costa. Ceci représente le nombre maximum de bits d'information qui peuvent être cachés dans un cover-document donné, lorsque la puissance d'insertion est infèrieure à la borne D_1 et sous une attaque type AWGN avec une puissance D_2.

Dans cette partie du travail, nous évaluons la capacité d'insertion pour chaque stégo-système ce qui revient à évaluer *le nombre maximum de bits d'information qui peuvent être insérés dans un document pour un système stéganographique donné [5].*

3.2.4 Robustesse (résistance)

Dans l'article [37], les auteurs définissent la robustesse dans un contexte de stéganographie avec gardien actif comme "*la quantification de la fiabilité du décodage en présence d'un canal bruité*". Evidemment, le bruit dans ce cas est ajouté par le gardien actif puisque le canal stéganographique est considéré comme sans bruit. Ainsi, la robustesse dans notre contexte est une évaluation de l'efficacité des attaques du gardien, puisque l'objectif de Wendy par ses attaques aveugles est d'augmenter la probabilité d'erreur du message transmis (même si celui-ci n'est pas détecté !) à la réception du côté de Bob.
Nous démontrons dans la suite de ce chapitre qu'il existe une sorte de proportionnalité entre la robustesse et la capacité (un même comportement : si l'une est croissante alors l'autre le sera forcément et inversement). Ceci n'est pas vrai pour tous les stégo-systèmes puisque la capacité est le nombre maximum de bits d'information qu'il est possible d'insérer et d'extraire si le sétgo-système est optimal. D'un autre côté, la robustesse mesure la fiabilité du décodeur utilisé par Bob. Par exemple, il est possible que la capacité d'insertion soit élevée mais -parce que le décodeur utilisé par Bob n'est pas optimal, par exemple- la robustesse du stégo-système est faible.

Dans la suite, nous procéderons à une analyse de quelques stégo-schémas informés. Nous utiliserons un ratio puissance du document hôte sur la puissance d'insertion compris dans l'intervalle $[0, 40]$ dB, lorsque le paramètre variable est la puissance d'insertion. Par contre, si la puissance d'attaque du gardien actif Wendy est prise comme paramètre variable pour les expériences, alors, nous considérons le ratio entre la puissance d'insertion et la puissance d'attaque compris dans l'intervale $[-20, 15]$ dB. Les intervalles considérées sont les plus proches des conditions réelles et permettent, ainsi, de comparer nos résultats avec ceux déjà existants dans le domaine

du data-hiding [1]. Dans cette partie du travail, nous proposons une étude détaillée des stégo-schémas basés sur les travaux de Costa connus pour leurs efficacités en terme de robustesse et de capacité.

3.3 Analyse du Schéma Scalaire de Costa (SCS)

Le SCS [1] est un système basé sur la quantification scalaire du cover-signal. Ce schéma découle des travaux de Costa sur le codage du canal avec information adjacente [14]. Considérons un message **m** à insérer et un cover-signal **s**. Pour le stégo-système SCS avec un stégo-message binaire, nous définissons deux dictionnaires utilisant deux quantificateurs décalés ("shiftés") :

$$\begin{aligned} \mathcal{U}_0[i] &= \{n\Delta + \mathbf{k}[i],\, n \in \mathbb{Z}\} \\ \text{et} & \\ \mathcal{U}_1[i] &= \{n\Delta + \mathbf{k}[i] + \frac{\Delta}{2},\, n \in \mathbb{Z}\}, \end{aligned} \tag{3.15}$$

où Δ est le pas de quantification et **k** représente la clef secrète [1]. Il est possible d'éviter l'utilisation de la stégo-clef **k** si Alice et Bob ne trouvent pas l'occasion de s'échanger une clef secrète avant d'être mis en prison (stéganographie à clef publique [45]). Dans ce cas, le message ne peut être crypté et il est plus difficile d'empêcher le gardien de lire le message secret (dans le cas d'un contexte malicieux [5], où le gardien tente de lire le message, qui est différent du contexte du contexte de ce chapitre). Rappelons quelques définitions données par Cox et al. [5] : "*la stego-clef est utilisé avec des algorithmes stéganographiques publiques pour insérer un message dans un cover-document.*
Comme dans le tatouage numérique, la clef peut contrôler, par exemple, les emplacements des modifications sur le cover-signal (le chemin d'insertion) ou peut servir à générer d'autres clefs secrètes plus longues ou encore à génerer d'autres paramètres qui rentrent dans le processus d'insertion...". Pour la suite, nous considérons la clef secrète tout paramètre qui est connu exclusivement par Alice et Bob (surtout pas Wendy !). Ces paramètres peuvent être, par exemple, le vecteur **k** avec des composantes considérées comme des variables pseudo-aléatoires continues utilisées dans

1. sans connaissance à priori de ce paramètre, il est très difficile de retrouver les dictionnaires et extraire le message caché.

le SCS (voir Eqn. (3.15)) ou une direction secrète pour le Spread Transform (ST). Notons que parfois la clef secrète est appelée "dithering" (en particulier pour les systèmes basés sur la quantification tel que le SCS), dans cette partie du travail, nous préférons utiliser le terme clef secrète ou stégo-clef.
Pour le schéma de tatouage SCS, d'après le bit $\mathbf{m}[i]$ à insérer, l'un des deux dictionnaires générés avec le quantificateur scalaire de pas Δ est choisi (i.e. $\mathcal{U}_0[i]$ si $\mathbf{m}[i] = 0$, et $\mathcal{U}_1[i]$ sinon), où l'élement du dictionnaire $\mathbf{u}^{\star}_{\mathbf{m},\mathbf{k}}[i]$ le plus proche de $\mathbf{s}[i]$ est choisi. Le stégo-signal est donné par l'expression suivante :

$$\mathbf{x}[i] = \mathbf{s}[i] + \alpha \left(\mathbf{u}^{\star}_{\mathbf{m},\mathbf{k}}[i] - \mathbf{s}[i] \right),$$

où $\alpha \in]0, 1]$ est le paramètre d'optimisation du schéma de Costa. Pour les expériences et les simulations, nous utilisons -par défaut- cette valeur optimale de α évaluée expérimentalement par Eggers et al. [1] : $\alpha = \sqrt{\sigma_W^2/(\sigma_W^2 + 2.71 \cdot \sigma_V^2)}$ -où σ_W^2 et σ_V^2 sont, respectivement, la puissance du tatouage et la puissance de l'attaque du gardien (canal)-. Par exemple, lorsque l'entropie relative $D(p_S||p_X)$ est calculée, la seule distorsion prise en compte est celle due à l'insertion : le canal est considéré non-bruité, ainsi, la valeur du paramètre α est égale à 1 (voir [1]). Autrement, la valeur du paramètre α sera spécifiée.
A l'encodeur, nous choisissons le mot de code le plus proche de la valeur de l'échantillon reçu $\mathbf{r}[i]$ dans l'union des deux dictionnaires (i.e. dans le cas binaire $\mathcal{U}_0[i] \cup \mathcal{U}_1[i]$). Pour cette partie du travail, nous considérons que le système SCS est notre stégo-système de référence, puisque tous les stégo-schémas étudiés dans ce travail sont basés sur ce système. Aussi, Eggers et al. ont démontré dans [1] que généralement la robustesse de SCS (et la capacité en particulier pour de faibles distorsions) contre les attaques est meilleure que celle du spread spectrum watermarking [26]. Dans la suite, nous proposons une analyse détaillée et une étude comparative entre les différentes possibilités d'améliorations du niveau d'indétectabilité et aussi les effets de chaque système sur le compromis entre les performances du stego-système.

Pour la suite des développements, $\mathbf{u}^{\star}_{\mathbf{m},\mathbf{k}}[i]$(ou $\mathbf{u}^{\star}_{\mathbf{m}}[i]$ lorsqu'aucune clef secrète n'est utilisée), $\mathbf{m}[i]$, $\mathbf{s}[i]$ et $\mathbf{x}[i]$ sont notés par $u_{m,k}$ (ou u_m), m, s et x pour rendre les résultats plus lisibles.

3.3.1 Densité de probabilité du stégo-signal

Pour un cover-signal représenté par un vecteur $\mathbf{s}$, un message $\mathbf{m}$ à insérer et le dictionnaire correspondant $\mathbf{u}^{\star}_{\mathbf{m}}$, la p.d.f. du stégo-signal $\mathbf{x}$ obtenu après un tatouage SCS sans clef secrète $\mathbf{k}$ est donnée par la formule suivante :

$$p_X(x) = \frac{1}{2(1-\alpha)} \sum_m \sum_{u_m} 1_{\left[u_m - \frac{(1-\alpha)\Delta}{2}, u_m + \frac{(1-\alpha)\Delta}{2}\right]}(x) \times p_S\left(\frac{x - \alpha u_m}{1-\alpha}\right), \tag{3.16}$$

où $1_{[.]}(.)$ représente la fonction fenêtre, α est le paramètre de Costa et Δ est le pas du quantificateur scalaire utilisé dans stégo-système SCS.

Preuve *Le cover et le stégo-signal sont considérés, respectivement, comme un ensemble de réalisations de variables aléatoires, indépendantes et non-stationnaires :* $\mathcal{S} = \{S[1], \ldots, S[N]\}$ *et* $\mathcal{X} = \{X[1], \ldots, X[N]\}$. *Lorsqu'un message* $\boldsymbol{m}$ *est inséré en utilisant le dictionnaire correspondant -modélisé par l'ensemble des réalisations aléatoires, indépendantes et non-stationnaires :* $\mathcal{U}_m = \{U_m[1], \ldots, U_m[N]\}$*-, le stégo-signal est donné par l'équation suivante (pour la suite, nous enlevons les index des variables pour rendre les equations plus lisibles, aussi, nous n'utilisons pas de clef secrète) :*

$$X = (1-\alpha)S + \alpha U_m, \tag{3.17}$$

où α *représente le paramètre de Costa [14]. D'après la règle du produit* $p(s, u_m|m) = p(s|u_m, m) \times p(u_m|m) = p(u_m|s, m) \times p(s|m)$, *et*

$$p(s|u_m, m) = \frac{p(u_m|s, m) p_S(s)}{p(u_m|m)},$$

nous avons :

$$p(u_m|s, m) = \begin{cases} 1 \ si \ s \in [u_m - \frac{\Delta}{2}, u_m + \frac{\Delta}{2}] \\ 0 \ autrement \end{cases}$$

ou

$$p(u_m|s, m) = \delta(u_m - Q_\Delta(s)), \tag{3.18}$$

où δ *représente le delta de Kronecker et* $Q_\Delta(.)$ *est le quantificateur scalaire avec un pas* Δ. *De plus, nous avons :* $\sum_{u_m} p(u_m|s, m) = 1$. *D'un autre côté,*

$$p(s|m) = \sum_{u_m} p(s|u_m, m) p(u_m|m) = \sum_{u_m} \delta(u_m - Q_\Delta(s)) p_S(s). \tag{3.19}$$

Si nous procédons au changement de variable suivant : $S = \frac{X-\alpha U_m}{1-\alpha}$*, dans la dernière équation, nous obtenons*

$$p(x|m) = \frac{1}{1-\alpha} \sum_{u_m} \delta\left(u_m - Q_\Delta\left(\frac{x-\alpha u_m}{1-\alpha}\right)\right) \times p_S\left(\frac{x-\alpha u_m}{1-\alpha}\right). \tag{3.20}$$

Lorsque les bits d'informations sont équiprobables, nous avons :

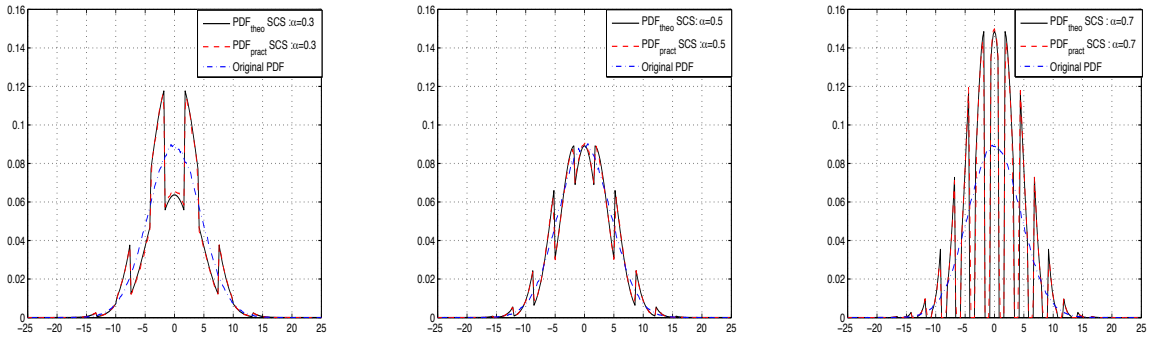
$$p_X(x) = \frac{1}{2(1-\alpha)} \sum_{u_m,m} \delta\left(u_m - Q_\Delta\left(\frac{x-\alpha u_m}{1-\alpha}\right)\right) \times p_S\left(\frac{x-\alpha u_m}{1-\alpha}\right). \tag{3.21}$$

Dans ce cas, la distance entre les points de reconstruction des deux quantificateurs est égale à $\Delta/2$, et toute fonction fenêtre se chevauche avec celle qui est la plus proche si $(1-\alpha)\Delta/2 > \Delta/4$ (ce qui est équivalent à $\alpha < 1/2$) ; et pour $\alpha > 1/2$ les fonctions indicatrices sont disjointes. Ceci explique les trous et les bosses sur la p.d.f. du stégo-signal sur les figure Fig. 4.11(a) et 4.11(c).

Pour $\alpha = 1/2$, à priori, il n'y a pas de trou ni de bosse sur la p.d.f. du stégo-signal mais nous obtenons une p.d.f. continue uniquement si $p_S(u/2) = p_S(u/2 + \Delta/4)$. La dernière égalité est satisfaite uniquement pour une densité uniforme du cover-signal. Pour une p.d.f. Gaussienne, plusieurs discontinuités apparaissent dans les points de liaisons, tel qu'il est montré sur la figure Fig. 4.11(b). Les discontinuités observées diminuent le niveau d'indétectabilité (augmentent le niveau de détectabilité), donc, le SS n'est pas un bon système dans un contexte stéganographique.
Dans l'article [2], les auteurs montrent que l'utilisation de la stégo-clef avec le stégo-système SCS permet d'améliorer le niveau d'indétectabilité tout en préservant la capacité et la robustesse du SCS.
Dans ce qui suit, nous étudions une amélioration du schéma basé sur le SCS proposé dans l'article [1], où les auteurs améliorent l'indétectabilité du système sans utiliser de clef secrète.

3.3.2 Amélioration du schéma scalaire de Costa (SCS) : schéma de Guillon et al.

En se basant sur les travaux d'Anderson et Petitcola [45], Guillon et al. [2] ont proposé un stégo-schéma. Fig. 3.3 résume les deux parties de ce schéma tel qu'il a



(a) (b) (c)

FIGURE 3.2 – Densité de probabilité du cover et du stégo-signal utilisant le stégo-système SCS pour $D_1 = 1$ et un cover-signal Gaussien de variance $\sigma_S^2 = 20$ avec différentes valeurs du paramètre de Costa α : (a) $\alpha = 0.3$, (b) $\alpha = 0.5$ and (c) $\alpha = 0.7$.

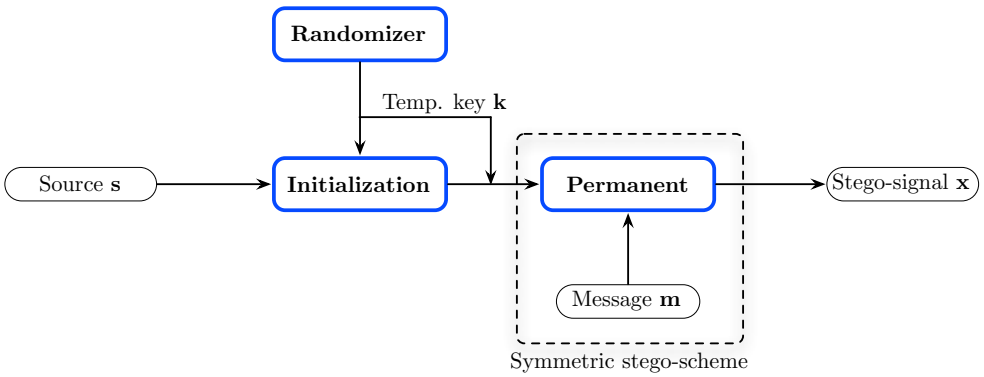


FIGURE 3.3 – Schéma de stéganographie asymétrique : la phase permanente est initialisée avec une clef privée temporaire **k**.

été proposé par les auteurs de l'article [2]. Dans la partie d'initialisation, une clef secrète **k** est générée avec un générateur pseudo-aléatoire et est cryptée avec un algorithme de cryptage asymétrique. La clef obtenue de **k** et de $\mathbf{k}_{\text{pub}}$ -où $\mathbf{k}_{\text{pub}}$ est une clef publique connue de tout utilisateurs- est insérée dans le cover-signal. La phase permanente utilise la clef transmise **k**, le stégo-schéma SCS et le message que l'on voudrait transmettre **m**.

Dans la phase permanente, l'indétectabilité du stégo-système SCS est principalement assurée par la clef privée transmise [2]. Cependant, la phase d'initialisation nécessite la transmission d'une information publique sans distordre le stégo-signal. Guillon et al. ont proposé d'utiliser le SCS avec un paramètre $\alpha = 1/2$ dans le but

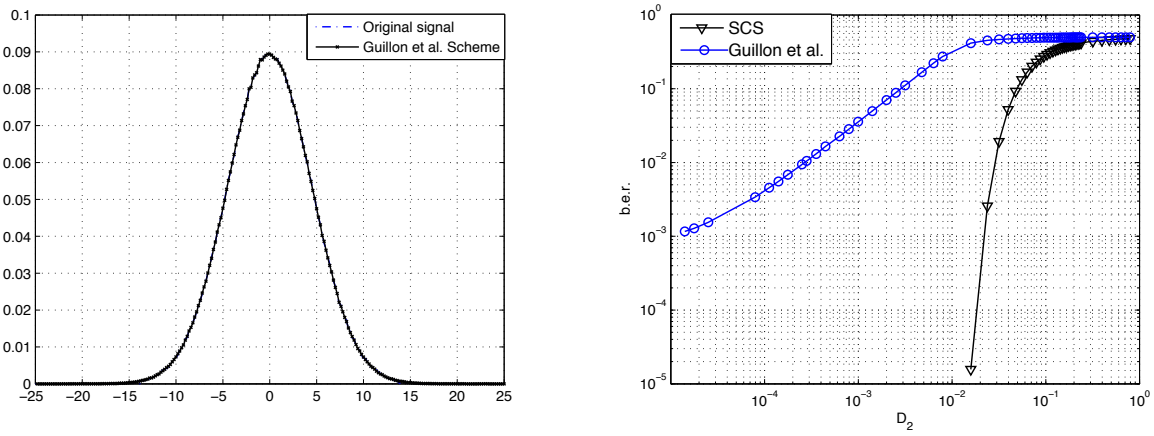
de cacher le message d'une manière indétectable, mais ceci n'est valable uniquement lorsque le cover-signal suit une loi uniforme ; ainsi, ils ont proposé d'utiliser un compresseur, qui travaille sur les histogrammes du signal marqué, avant l'insertion de l'information dans le but d'égaliser la p.d.f. du cover-signal. Le message inséré est donc statistiquement invisible (voir [2] pour plus de justifications) tel qu'il est montré sur Fig. 3.4(a). Malheureusement, le stégo-système résultant est moins flexible parce que les étapes d'encodage et de décodage dépendent fortement des statistiques du cover-signal et le récepteur aussi nécessite une connaissance à priori de ces statistiques. D'un autre côté, tel qu'il est décrit par la figure Fig.3.4 (b), le schéma est moins résistant face aux attaques du gardien actif par rapport au schéma classique SCS. Ceci est dû essentiellement à la compression et décompression qui entraînent des distorsions importantes.
Il a été montré [46] que les artéfacts dans le stégo-signal sont essentiellement dus à l'utilisation de dictionnaires à partitionnement régulier obtenue à l'aide de quantificateur scalaires uniformes. Dans la section suivante, nous proposons d'utiliser un dictionnaire non-uniforme et très structuré basé sur la TCQ proposé à la base dans le codage de source [30].

3.4 Analyse du stégo-schéma basé sur la TCQ

L'approche proposée dans cette section concerne l'utilisation de la quantification basée sur le treillis, pour un partitionnement pseudo-aléatoire des dictionnaires, dans le but d'éliminer les artéfacts introduits dans la p.d.f. du stégo-signal en évitant un partitionnement uniforme (comme il a été fait avec le stégo-schéma SCS).

3.4.1 Principe de base de la TCQ

En se basant sur les travaux de Cox et al. [5] et de la même façon que le papier de Gaëtan Le Guelvouit [46], nous allons décrire une implémentation pratique de la TCQ appliquée à stéganographie.
Considérons le treillis définit par la fonction de transition : $\mathcal{E} \times \{0,1\} \longrightarrow \mathcal{E}$, tr : $(\mathbf{e}[i], \mathbf{m}[i]) \longmapsto \mathbf{e}[i+1]$, tel que : $\mathcal{E} = \{0, 1, \ldots, 2^{r-1}\}$ représente l'ensemble des états du treillis possibles, où r est un entier supérieur à 1 : $r > 1$, et i est l'index de la transition courante (index temporel). Contrairement au SCS, les mots de code $\mathbf{u}_{\mathbf{m}}[i]$ ne dépendent pas uniquement du message inséré mais devient aussi fonction de l'état



(a) (b)

FIGURE 3.4 – (a) Fonctions densités de probabilité d'un cover-signal Gaussien, de variance $\sigma_S^2 = 20$, et du stego-signal pour le schéma de Guillon *et al.* [2] dont la puissance d'insertion est égale à 1. (b) Taux d'erreur binaire (b.e.r.) induit par les attaques du gardien Wendy de puissance D_2 dans le cas du stégo-système SCS and la version améliorée du SCS proposée par le schéma de Guillon *et al.*, telle que la puissance d'insertion $D_1 = 1$ (La variance du cover-signal Gaussien σ_S^2 est égale à 20).

actuel du treillis, ainsi, le décalage $\mathbf{d}[i]$ (dans le stégo-système SCS ce décallage est égal à : $m \cdot \Delta/2$ tel qu'il est montré dans Eqn.3.15) peut être exprimé comme suit (voir [46] pour plus de détails) :

$$\mathcal{E} \times \{0,1\} \longrightarrow [-\Delta/2, +\Delta/2], \tag{3.22}$$

$$f : (\mathbf{e}[i], \mathbf{m}[i]) \longmapsto \mathbf{d}[i]. \tag{3.23}$$

Dans ce stégo-système, les dictionnaires sont définit par :

$$\mathcal{U}_{\mathbf{m}}[i] = \{n\Delta + f(\mathbf{e}[i], \mathbf{m}[i]),\, n \in \mathbb{Z}\},$$

et le mot de code $\mathbf{u}^{\star}_{\mathbf{m}}[i]$ le plus proche de $\mathbf{s}[i]$ est choisit de manière à ce que ce mot de code appartienne à l'ensemble $\mathcal{U}_{\mathbf{m}}[i]$:

$$\mathbf{u}^{\star}_{\mathbf{m}}[i] = \arg \min_{\mathbf{u} \in \mathcal{U}_{\mathbf{m}}[i]} \sum_{j=1}^{G} (\mathbf{s}[i] - \mathbf{u}[j])^2, \tag{3.24}$$

où $G \in \mathbb{N}^{\star}$ le nombre d'état du treillis. Pour les expériences utilisant le stégo-système TCQ, nous prenons le nombre d'états du treillis égal à : 2^9 sauf indication contraire. Notons que l'index j est différent de l'index i. j représente l'index du vecteur d'états du treillis $\mathbf{e}_t$, le vecteur regroupant tout les états possibles du treillis utilisé dans stégo-système basé sur la TCQ. Par contre, l'index i représente l'index temporel de la transition courante.
Le stégo-signal est donné par :

$$\mathbf{x}[i] = \mathbf{s}[i] + \alpha\,(\mathbf{u}^{\star}_{\mathbf{m}}[i] - \mathbf{s}[i]), \tag{3.25}$$

comme pour le stégo-schéma SCS, $\alpha \in]0, 1]$ représente le paramètre d'optimisation du schéma Costa, aussi, pour les expériences et les simulations, nous utilisons la valeur optimale du paramètre α sauf indication contraire. Notons que la valeur optimale du paramètre α est un peu différente entre le stégo-système TCQ et SCS. Pour le SCS, Eggers et al. [1] ont développé une formule expérimentale pour calculer le paramètre α, par contre, les expériences sur le stégo-système TCQ montrent que la valeur optimale de ce paramètre coïncide avec celle donnée par la formule de Costa dans son fameux article "*Writting on the dirty paper*" [14] : $\alpha = \sigma_w^2/(\sigma_w^2 + \sigma_v^2)$ (σ_w^2 and σ_v^2 sont respectivement la puissance d'insertion et de l'attaque du warden sur le

stégo-signal). Pour extraire le message l'algorithme de Viterbi [32] est utilisé dans le but de retrouver le chemin correspondant au stégo-message.

3.4.2 Analyse des performances

Théorème 1 : Pour un cover-signal et un signal d'insertion modélisés respectivement par la variable aléatoire S et E de variance σ_S^2 et σ_E^2. Lorsque le stégo-système TCQ avec un total de nombre d'états du treillis important, la p.d.f. du stégo-signal modélisé par la variable aléatoire X. Pour $X = x$ est la moyenne du cover-signal sur l'intervalle $\left[x - \sigma_E\sqrt(3), x + \sigma_E\sqrt(3)\right]$, donc,

$$p_X(x) = \frac{1}{\sigma_E\sqrt{12}} \int_{x-\sigma_E\sqrt{3}}^{x+\sigma_E\sqrt{3}} p_S(z)\ \mathrm{d}z. \tag{3.26}$$

Lorsque la variance σ_E^2 est petite par rapport à la variance du cover-signal σ_S^2, la p.d.f. p_X pour $X = x$ converge vers p_S pour $S = x$.

Preuve : *Nous reprenons les modèles et les hypothèses émises sur le stégo et le cover-signal dans la section 3.3.1. Les états du treillis sont notés par* $\boldsymbol{e_t}[j]$ *– pour* $j = 1, \ldots, G$ *–, aussi, nous supposons que les valeurs des états du treillis suivent une distribution uniforme* $p_{E_t}(e_t) = 1/G$*. Nous nous basons sur l'implémentation pratique du stégo-système TCQ proposée dans l'article [30]. L'affectation des dictionnaires pour chaque transition du treillis se fait en décalant une version de base d'un quantificateur scalaire de pas* Δ *:* $Q_\Delta(.)$ *(Similaire à la construction des dictionnaires du stégo-système SCS dont le nombre est limité à deux seulement). Donc, les échantillons hôtes seront remplacés par* $U_{(n,m,e_t)}$*,* $n \in \mathbb{Z}$*, un mot de code des sous-dictionnaires, obtenus grâce à la quantification scalaire qui correspond à l'état* e_t *et le bit message* m*. Une analyse du stégo-système TCQ permet de dévolopper une formule générale des mots de codes choisis dans le dictionnaire correspondant au chemin du treillis définit par le message inséré. Ainsi, les mots de code sont formulés comme :* $U_{(n,m,\boldsymbol{e_t}[j])} = (n + m/2 - j/G)\Delta$ *pour* $j = 1, \ldots, G/2$ *et* $U_{(n,m,\ \boldsymbol{e_t}[j])} = U_{n,m,\ \boldsymbol{e_t}[j-G/2]}$ *pour* $j = G/2+1, \ldots, G$*. En utilisant les résultats du développement précédent -pour la p.d.f. d'un signal marqué avec le SCS-, la formule de la p.d.f. d'un stégo-signal tatoué avec le système TCQ -sachant un état fixé* e_t*- est donnée par :*

$$p(x|e_t) = \frac{1}{2\,(1-\alpha)} \sum_{n,m} 1_{[-\frac{1}{2(1-\alpha)}, \frac{1}{2(1-\alpha)}]}(x - u_{(n,m,e_t)}) p_S\left(\frac{x - \alpha u_{(n,m,e_t)}}{1-\alpha}\right).$$

En procédant à une marginalisation sur tous les états possibles $\boldsymbol{e_t}[j], j = 1, ..., G$, *nous obtenons*

$$\begin{aligned} p_X(x) &= \sum_{j=1}^{G} p_X(x|\boldsymbol{e_t}[j]) p_E(\boldsymbol{e_t}[j]) \\ &= \frac{1}{(1-\alpha)} \sum_{n,m} \frac{1}{G} \sum_{j=1}^{G/2} 1_{[-\frac{1}{2(1-\alpha)}, \frac{1}{2(1-\alpha)}]} \left(x - u_{(n,m,\boldsymbol{e_t}[j])}\right) \times p_S \left(\frac{x - \alpha u_{(n,m,\boldsymbol{e_t}[j])}}{1-\alpha} \right) \\ &= \frac{1}{(1-\alpha)} \sum_{n,m} \frac{1}{2} \times \frac{1}{G/2} \sum_{j=1}^{G/2} 1_{[-\frac{1}{2(1-\alpha)}, \frac{1}{2(1-\alpha)}]} \left(x - \left(n + \frac{m}{2} - \frac{j}{G/2} \times \frac{1}{2} \right) \Delta \right) \\ &\quad \times p_S \left(\frac{x - \alpha \left(n + \frac{m}{2} - \frac{j}{G/2} \times \frac{1}{2} \right) \Delta}{1-\alpha} \right). \end{aligned} \tag{3.27}$$

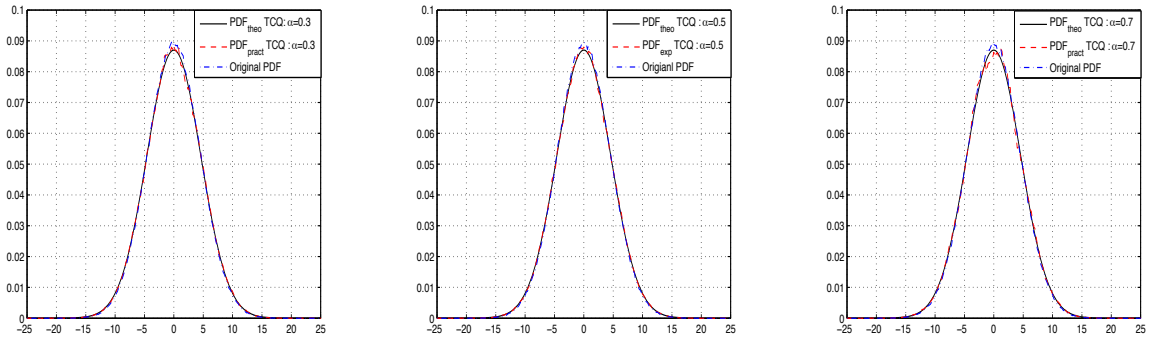
Lorsque le nombre d'états du treillis est très grand et en utilisant les propriétés de la somme de Riemann, nous aurons :

$$p_X(x) = \frac{1}{1-\alpha} \sum_{n,m} \int_0^{\frac{1}{2}} 1_{[-\frac{1}{2(1-\alpha)}, \frac{1}{2(1-\alpha)}]} \left(x - (n + \frac{m}{2} - \gamma) \Delta \right) \times p_S \left(\frac{x - \alpha \left(n + \frac{m}{2} - \gamma \right) \Delta}{1-\alpha} \right) \, d\gamma. \tag{3.28}$$

Si nous remplaçons m par ces deux valeurs possibles, i.e. 0 ou 1, et en appliquant le changement de variable :$Z = \frac{X - \alpha\gamma\Delta}{1-\alpha}$, *nous obtenons*

$$p_X(x) = \frac{1}{\alpha\Delta} \int_{x - \frac{\alpha\Delta}{2}}^{x + \frac{\alpha\Delta}{2}} p_S(z) \, \mathrm{d}z = \frac{1}{\sigma_E \sqrt{12}} \int_{x - \sigma_E\sqrt{3}}^{x + \sigma_E\sqrt{3}} p_S(z) \, \mathrm{d}z. \tag{3.29}$$

Nous avons implémenté Eqn.3.26 pour un signal de densité de probabilité Gaussienne et nous obtenons les résultats présentés sur la Fig.3.5. On peut noter que les deux p.d.f. estimées des signaux marqués sont quasiment les mêmes lorsqu'on les calcule expérimentalement et à l'aide de l'expression théorique. De plus, Fig.3.5 prouve que le paramètre α n'affecte pas la p.d.f. du stégo-signal (le comportement de la p.d.f. du signal tatoué à l'aide de la TCQ est indépendante du paramètre α), tel qu'il est montré par Eqn.3.26, même si les conditions théoriques ne sont pas toutes complètement vérifiées à cause de contraintes pratiques. Notons que la densité de probabilité du stégo-signal est légèrement différente de celle du cover-signal. Ceci ne veut pas dire que la TCQ est un stégo-ssystème qui ne vérifie pas la contrainte



(a) (b) (c)

FIGURE 3.5 – Fonctions densités de probabilité d'un cover-signal Gaussien, de variance $\sigma_S^2 = 20$, et d'un stégo-signal pour une puissance d'insertion $D_1 = 1$ en utilisant le stégo-système TCQ pour différentes valeurs du paramètre α : (a) $\alpha = 0.3$, (b) $\alpha = 0.5$ and (c) $\alpha = 0.7$.

d'indétectabilité, puisque cette petite différence est due à la puissance d'insertion importante (le ratio entre la puissance du cover-signal et la puissance d'insertion est égal à 13 dB). Ce choix est dicté par notre volonté de voir les limites de notre système. Cependant, des expériences similaires ont été effectuées avec des puissances d'insertions modérées qui prouvent que la TCQ est un stégo-système indétectable. Dans Fig.3.6(a), où un signal Gaussien de variance $\sigma_S^2 = 20$ est utilisé, et Fig.3.6(b), où 100 images réelles de taille 350×350 pixels sont utilisées, valident le modèle théorique formulé dans Eqn.3.2 (le fait que $D_{theo}(p_S||p_X)$ est une fonction strictement décroissante par rapport à la puissance d'insertion D_1). Les distorsions -induites par l'insertion TCQ du stégo-message- dans la p.d.f. du stégo-signal sont très limitées (l'entropie relative $D(p_S||p_X)$ est très petite) dans le cas où la puissance d'insertion a une valeur modérée par rapport à la variance σ_S^2 du cover-signal.

Les figures dans Fig.3.6 donnent l'entropie relative entre les p.d.f. des cover et stégo-signaux. Notons que le stégo-système TCQ permet une meilleure indétectabilité que le stégo-schéma SCS. Contrairement au schéma de Guillon et al. le schéma TCQ permet d'obtenir un système fléxible (puisque le système est complètement indépendant du cover-signal). De plus, le message inséré résiste aux différentes distorsions tout en préservant l'indétectabilité tel qu'il est montré par le compromis Robustesse-indétectabilité donné sur la figure 3.7(a). Fig.3.7(c) montre que le com-

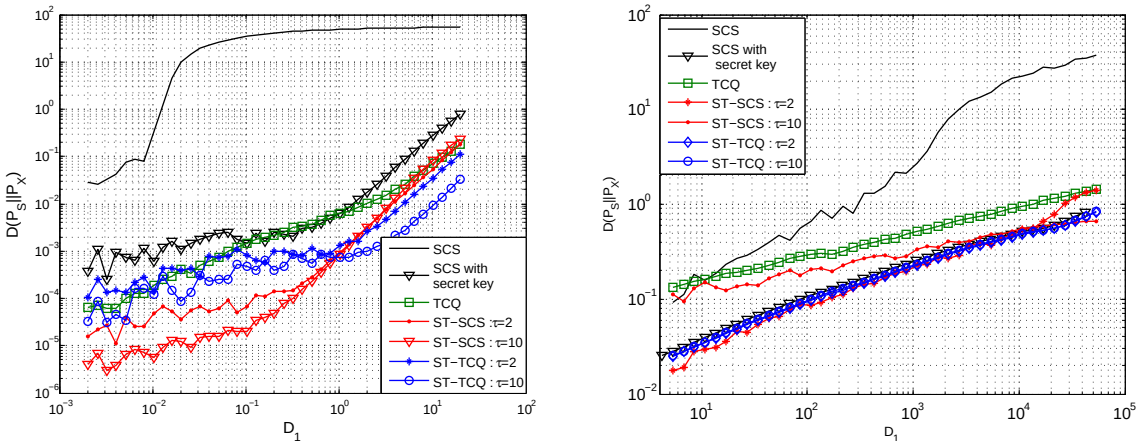


FIGURE 3.6 – (a) Entropie relative 1-Dimension entre les p.d.f. du stégo-signal (signal Gaussien de variance $\sigma_S^2 = 20$) et du cover-signal en fonction de la puissance d'insertion D_1, dans le cas des stégo-schémas SCS, TCQ, ST-SCS et ST-TCQ. (b) Entropie relative 1-Dimension entre les p.d.f. des cover and des stego-images réelles (nous utilisons 100 images réelles différentes) de taille 350×350 pixels.

pormis capacité-indétectabilité de la TCQ est meilleur que celui du schéma SCS lorsqu'aucune clef secrète n'est utilisée (parfois appelée dithering). Sur les Fig.3.7(b) et Fig.3.7(d), nous montrons que la capacité et la résistance contre le gardien actif du schéma TCQ ne sont pas aussi bonnes que celles du SCS, particulièrement, pour des attaques puissantes du gardien. Donc, même si le système TCQ offre un très bon niveau d'indétectabilité, il n'est pas le plus efficace dans un contexte gardien actif. Par exemple, pour un certain niveau de distorsion, nous montrons les perfomances du schéma TCQ pour les différentes propriétés (voir Fig.3.7 et suivre le rectangle à trait continue).

Dans le but de proposer un stégo-système avec de très bonnes performances dans un contexte de stéganographie avec gardien actif, nous allons étudier dans la section suivante la combinaison des stégo-systèmes avec le Spread Transform connu pour sa bonne robustesse.

3.5 Le spread Transform (ST)

Les schémas résultant de la combinaisons du ST avec les schéma de tatouage basés sur la quantifications est connue pour être un moyen très efficace pour amé-

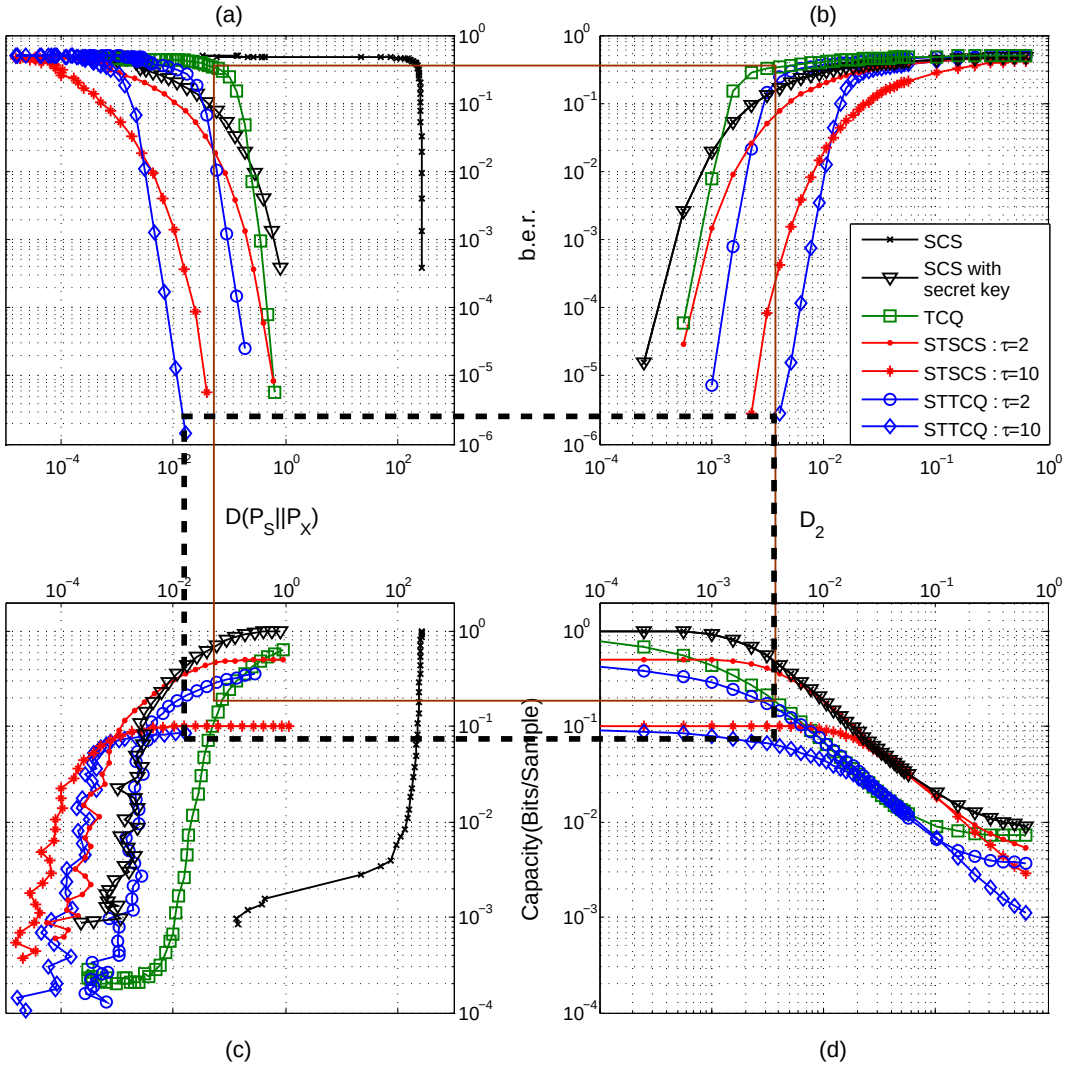


FIGURE 3.7 – Performances du SCS, TCQ, ST-SCS and ST-TCQ stego-systeme avec un cover-signal Gaussien de variance $\sigma_S^2 = 20$: (a) BER vs. entropie relative, (b) BER en fonction de la puissance de l'attaque du gardien actif D_2, (c) capacité vs. entropie relative, (d) capacité en fonction dela puissance d'attaque du gardien actif D_2.

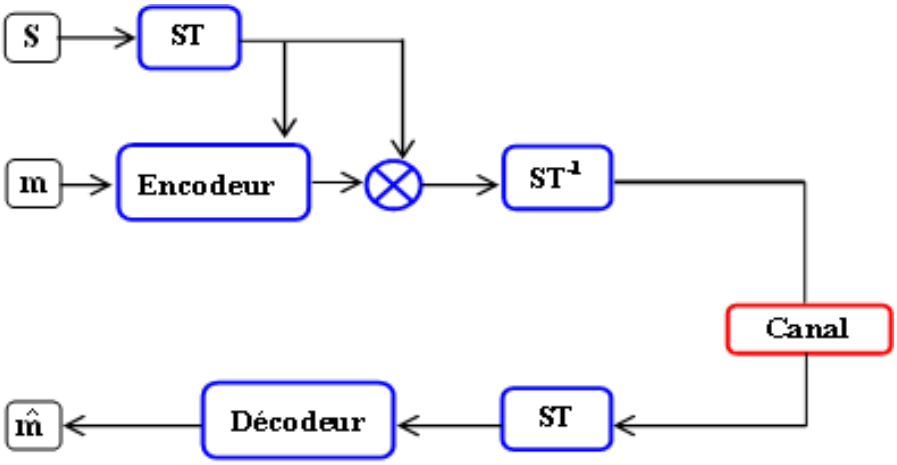


FIGURE 3.8 – Spread transform combiné avec des stégo-systèmes informés.

liorer la robustesse face aux attaques AWGN [1]. Un rappel très rapide du Spread Transform est donné dans la suite.

Chen et Wornel dans [15] ont introduit un schéma de tatouage très efficace qui permet d'étaler l'information dissimulée sur plusieurs échantillons hôtes. Dans ce chapitre, nous utilisons cette technique dans le contexte spécifique de la stéganographie avec gardien actif dans le but d'améliorer la résistance de nos stégo-systèmes face aux attaques du gardien. Le processus global est décrit sur Fig.3.8.

La transformation du cover-signal $\mathbf{s}$ de taille
$N \in \mathbb{N}^\star$, noté par $\mathbf{s}^{\text{st}} = [s^{\text{st}}[0], \ldots, s^{\text{st}}[N/\tau - 1]]$, est donnée par

$$\mathbf{s}^{\text{st}}[l] = \sum_{i=\tau l}^{\tau l+\tau-1} \mathbf{s}[i] \times \mathbf{t}[i], \tag{3.30}$$

où $\tau \in \mathbb{N}^\star$ est le facteur d'étalement. Cette opération est plutôt une projection selon une direction $\mathbf{t}$, où $\mathbf{t}$ est un vecteur normalisé. Dans la suite de ce chapitre, $\mathbf{t}[i]$ est noté par t pour plus de clareté des equations. Puis, une transformation inverse est appliquée au stégo-signal transformé : $\mathbf{x}^{\text{st}}[i] = \mathbf{s}^{\text{st}}[i] + \mathbf{e}^{\text{st}}[l]$, tel que $\mathbf{e}^{\text{st}}[l]$ est le l^{me} échantillon du signal d'insertion dans le domaine transformé $\mathbf{e}^{\text{st}}$ et nous avons

$$\mathbf{x}[i] = \mathbf{s}[i] + \underbrace{\mathbf{e}^{\text{st}}[l] \times \mathbf{t}[i]}_{\mathbf{e}[i]}, \tag{3.31}$$

tel que $i \in \{0, \ldots, N-1\}$ and $l \in \{0, \ldots, N/\tau - 1\}$. Notons que Eqn.(3.31), $\mathbf{e}^{st}$ est générée par l'encodeur dans le domaine transformé.
Tel qu'il a été signalé auparavant, les attaques du gardien actif sont modélisées par une attaque AWGN lors de la transmision à travers le canal et finalement l'encodeur recevrait $\mathbf{y}[i] = \mathbf{s}[i] + \mathbf{e}^{\mathrm{st}}[l] \times \mathbf{t}[i] + \mathbf{v}[i]$. Avant le décodage du message $\mathbf{m}$, le signal reçu $\mathbf{y}$ est transformé, ceci implique $\mathbf{y}^{\mathrm{st}}[l] = \mathbf{s}^{\mathrm{st}}[l] + \mathbf{e}^{\mathrm{st}}[l] + \mathbf{v}^{\mathrm{st}}[l]$.

Le bon niveau de robustesse (équivalent à la bonne résistance face aux attaques du gardien actif) offert par le ST est expliqué par l'expression suivante :

$$\sigma_E^2 = \mathbb{E}\left[\left(E^{\mathrm{st}} \cdot T\right)^2\right] = \mathbb{E}\left[\frac{1}{\tau}\left(E^{\mathrm{st}}\right)^2\right] = \frac{\sigma_{E^{\mathrm{st}}}^2}{\tau} \Rightarrow \sigma_{E^{\mathrm{st}}}^2 = \tau\sigma_E^2, \tag{3.32}$$

où le signal d'insertion dans le domaine transformé E^{st} a une moyenne nulle et $\mathbb{E}[.]$ représente l'opérateur espérance mathématique d'une variable aléatoire. Ainsi, l'utilisation du ST permet de multiplier la puissance d'insertion par un facteur τ juste avant l'extraction du message. D'un autre côté, le ST permet d'améliorer la fidélité de tout système de dissimulation sans perdre en robustesse, puisque la transformation inverse permet de diviser la puissance d'insertion d'un facteur τ. Dans la suite, nous considérons un paramètre d'étalement $\mathbf{t}$ où : $\mathbf{t}[i] = \pm 1/\sqrt{\tau}$ (ce choix particulier nous permet d'étaler les distorsions dues à l'insertion de l'information uniformement sur tous les échantillons marqués). Aussi, la puissance d'insertion D_1 utilisée dans les expériences est celle du tatouage dans le domaine non-transformé.

Analyse des pérfomances du ST-SCS

Dans cette partie, nous procédons à une analyse théorique et expérimentale du stégo-système ST-SCS [1] dans le but d'évaluer l'indétectabilité.

Théorème 2 : Si S est une variable aléatoire modélisant le cover-sinal et U_m le codeword correspondant au stégo-message m, alors, la fonction densité de probabilité du stégo-signal, tatouée à l'aide le ST-SCS avec un facteur d'étalement τ et une direction d'étalement donnée par le vecteur $\mathbf{t}$ avec des composantes représentées par le scalaire t qui peuvent prendre deux valeurs possibles $\pm 1/\sqrt{\tau}$, est donnée par

$$
\begin{aligned}
p_X(x) &= \frac{\tau}{4(\tau-\alpha)} \\
&\sum_m \sum_{u_m t} \int_{-\infty}^{\infty} \delta\left(u_m - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right) t + y\right)\right) \\
&\times p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right)\right) p_Y(y)\, \mathrm{d}y, \qquad (3.33)
\end{aligned}
$$

où y est une variable auxiliaire, α est le paramètre de Costa et Δ représente le pas du quantificateur scalaire utilisé dans le système ST-SCS.

D'un autre côté, si nous remplacions t avec ces deux réalisation possibles, i.e., $\pm 1/\sqrt{\tau}$, et si nous prenions $\tau \to \infty$ avec une variance σ_S^2 du cover-signal $\mathbf{s}$, le stégo-signal $\mathbf{x}$ aurait la même fonction densité de probabilité que son cover-signal $\mathbf{s}$.

Preuve :

Dans ce développement, nous considérons le même modêle et les mêmes hypothèses concernant le stégo-système que dans la preuve du *Théorème 1.* La transformation du signal hôte $\mathbf{s}$ est modélisée par l'ensemble des réalisations de variables aléatoires, indépendantes et non-stationnaires, i.e. $\mathcal{S}^{\mathrm{st}} = \{S^{\mathrm{st}}[1], \ldots, S^{\mathrm{st}}[N/\tau]\}$. Aussi, nous considérons la direction d'étalement $\mathbf{t}$. Elle est modélisée par l'ensemble de variables aléatoires discrètes, indépendantes et de distribution uniforme, i.e. $\mathcal{T} = \{T[1], \ldots, T[N]\}$. Donc, lorsque le ST-SCS est utilisé pour l'insertion du message, le signal marqué X est donné par $X = S + \alpha(U - S^{\mathrm{st}}) \times T$, si nous considerions :

$$
S^{\mathrm{st}}[l] = \sum_{i=\tau l}^{\tau l+\tau-1} S[i] \times T[i] = S[n] \times T[n] + \underbrace{\sum_{i \neq n} S[i] \times T[i]}_{Y_n[l]} = S[n] \times T[n] + Y_n[l], \quad (3.34)
$$

où Y est consiérée comme une variable aléatoire modélisée par l'ensemble $\mathcal{Y} = \{Y_1[1], \ldots, Y_N[N/\tau]\}$. Notons que $S[n]$ est indépendant de $T[n]$. D'après la définition de $Y_n[l]$ dans Eqn. (3.34), $S[n]$ est aussi indépendante de $Y_n[l]$. Dans ce qui suit, nous enlevons l'index des variables aléatoires pour rendre les équations plus lisibles.

$$
X = S + \alpha(U_m - S \times T - Y) \times T. \qquad (3.35)
$$

Puisque $\forall i,\ \mathbf{t}^2[i] = t^2 = 1/\tau$, $T^2 = 1/\tau$, l'équation précédente devient donc,

$$X = \left(1 - \frac{\alpha}{\tau}\right) \times S - \alpha Y \times T + \alpha U_m \times T. \tag{3.36}$$

La p.d.f. du mot de code U_m conditionnellement à S, Y, T et le message m est donnée par,

$$p(u_m|s, y, t, m) = \delta\left(u_m - Q_\Delta\left(s \times t + y\right)\right). \tag{3.37}$$

Alors,

$$p(s|u_m, y, t, m) = \frac{\delta\left(u_m - Q_\Delta\left(s \times t + y\right)\right) p(s|y, t, m)}{p(u_m|y, t, m)}. \tag{3.38}$$

Dans cette partie, la variable aléatoire S est considérée comme variable indépendante de T et Y. Ainsi,

$$p(s|y, t, m) = p(s)$$

et

$$p(s|u_m, y, t, m) = \frac{\delta\left(u_m - Q_\Delta\left(s \times t + y\right)\right) p_S(s)}{p(u_m|y, t, m)}. \tag{3.39}$$

En procédant au changement de variable suivant,

$$S = \frac{\tau}{\tau - \alpha} \times (X + \alpha Y \times T - \alpha U_m \times T), \tag{3.40}$$

nous obtenons donc,

$$\begin{aligned} p(x|u_m, y, t, m) &= \frac{\tau}{(\tau - \alpha)} \frac{\delta\left(u_m - Q_\Delta\left(\frac{\tau}{(\tau-\alpha)}\left(x + \alpha y \times t - \alpha u_m \times t\right)\right) \times t + y\right)}{p(u_m|y, t, m)} \\ &\quad \times p_S\left(\frac{\tau}{\tau - \alpha}\left(x + y - \alpha u_m \times t\right)\right), \end{aligned} \tag{3.41}$$

en mariginalisant sur U_m, nous aurons,

$$\begin{aligned} p(x|y,t,m) &= \sum_{u_m} p(u_m|y,t,m) \times p(x|u_m,y,t,m) \\ &= \frac{\tau}{(\tau-\alpha)} \sum_{u_m} \delta\left(u_m - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right) \times t + y\right)\right) \\ &\quad \times p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right)\right). \end{aligned} \tag{3.42}$$

Puisque T est une variable aléatoire avec des réalisations prenant deux valeurs possibles $\pm 1/\sqrt{\tau}$ et que m est considérée comme équiprobable, la marginalisation sur ces deux variables mène à la formulation suivante,

$$\begin{aligned} p(x|y) &= \frac{\tau}{4\,(\tau-\alpha)} \sum_{m} \sum_{u_m,t} \delta\left(u - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right) \times t + y\right)\right) \\ &\quad \times p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right)\right), \end{aligned} \tag{3.43}$$

donc,

$$\begin{aligned} p_X(x) &= \frac{\tau}{4\,(\tau-\alpha)} \\ &\quad \sum_{m} \sum_{u_m,t} \int_{-\infty}^{\infty} \delta\left(u_m - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right) \times t + y\right)\right) \\ &\quad \times p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y \times t - \alpha u_m \times t\right)\right) p_Y(y)\ \mathrm{d}y. \end{aligned} \tag{3.44}$$

Dans le cas limite, lorsque le facteur d'étalement τ tend vers l'infini : $\tau \to \infty$, Eqn.3.44 devient,

$$\begin{aligned} p_X(x) &= \frac{1}{2} \sum_{m} \sum_{u_m} \int_{-\infty}^{\infty} \delta\left(u_m - Q_\Delta\left(y\right)\right) \times p_S(x) p_Y(y)\ \mathrm{d}y \\ &= \frac{1}{2} \sum_{m} \sum_{u_m} \int_{u_m-\Delta/2}^{u_m+\Delta/2} p_S\left(x\right) p_Y(y)\ \mathrm{d}y, \\ &= \int_{-\infty}^{\infty} p_S\left(x\right) p_Y(y)\ \mathrm{d}y \;=\; p_S\left(x\right) \int_{-\infty}^{\infty} p_Y(y)\ \mathrm{d}y \;=\; p_S\left(x\right). \end{aligned} \tag{3.45}$$

Sur Fig.3.9, la p.d.f. expérimentale du stégo-signal valide le modèle théorique

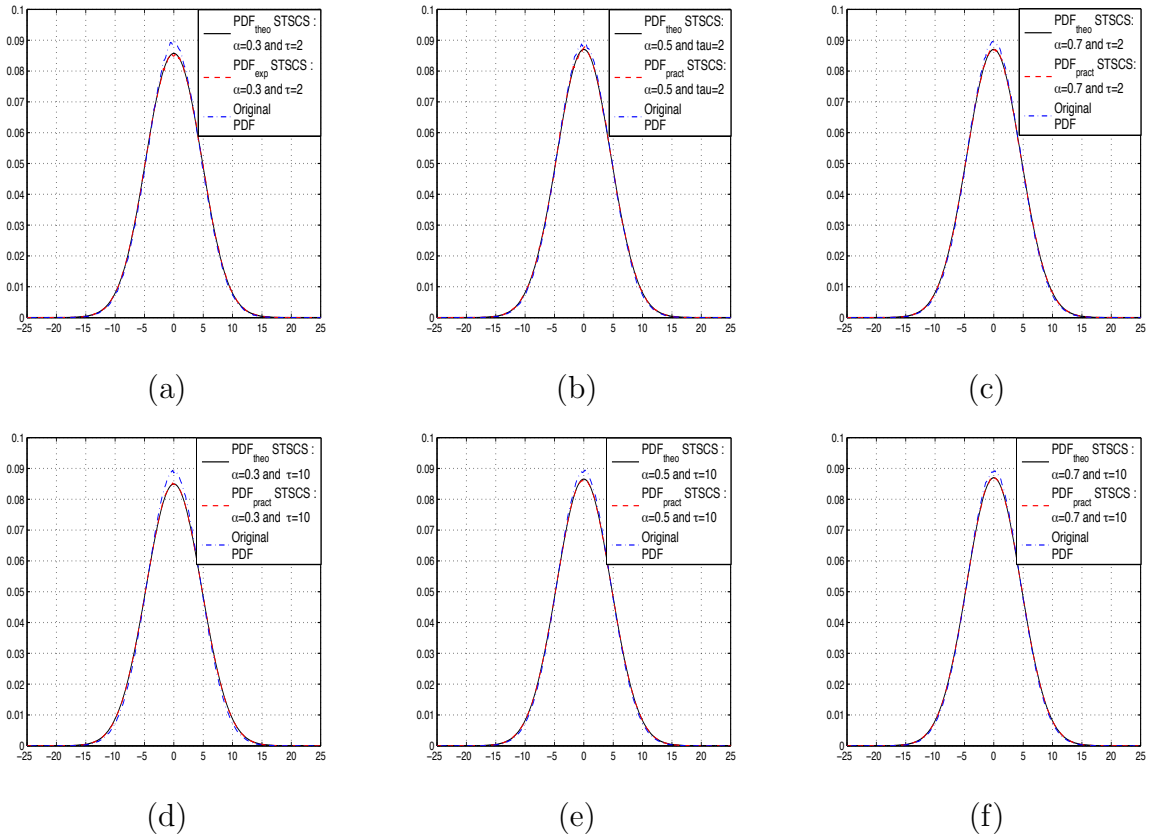
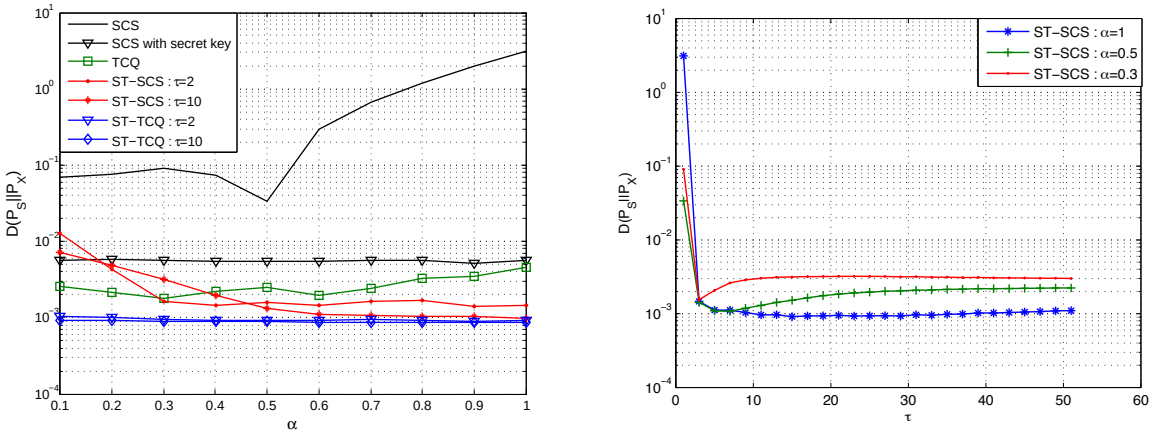


FIGURE 3.9 – Fonctions densités de probabilité d'un cover-signal Gaussien, de variance $\sigma_S^2 = 20$, et du stego-signal en utilisant le stégo-schéma ST-SCS, avec une puissance d'insertion $D_1 = 1$, pour $\tau = 2$ avec des différente valeurs du paramètre α : (a) $\alpha = 0.3$, (b) $\alpha = 0.5$ and (c) $\alpha = 0.7$; et pour $\tau = 10$ avec (d) $\alpha = 0.3$, (e) $\alpha = 0.5$ et (f) $\alpha = 0.7$.

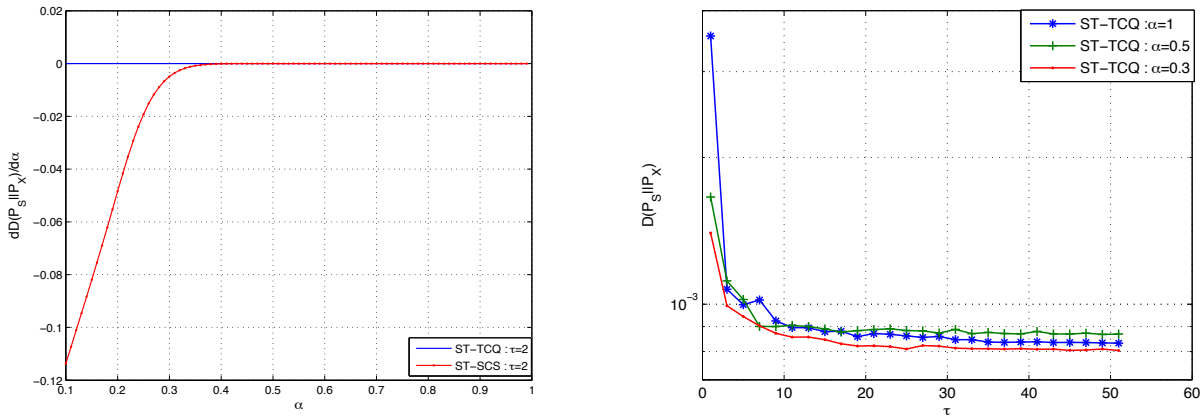


(a) (b)

FIGURE 3.10 – L'entropie relative entre un cover-signal Gaussien de variance $\sigma_S^2 = 20$ et du stego-signal où la puissance d'insertion D_1 est égale à 1 en fonction du (a) paramètre α avec différentes valeurs du facteur d'étalement τ pour les stégo-systèmes SCS, TCQ, ST-SCS, ST-TCQ ; et (b) le facteur d'étalement τ pour différentes valeurs de α avec le stégo-système ST-SCS.

donné par Eqn.3.33, ainsi, il nous est possible de dire que les résultats donnés par le modèle théorique suivent bien ceux obtenus expérimentalement.

Dans le cas limite (le facteur d'étalement τ considéré comme très grand), nous pouvons dire que le ST-SCS préserve parfaitement la p.d.f. du stégo-signal (voir Eqn. 3.45 dans la preuve du *Théorème 2*). Dans Fig.3.9, nous montrons que les distorsions induites par le tatouage ST-SCS sont très limitées, y compris, dans un cas proche de la réalité où le facteur d'étalement n'est pas infini. Ceci est aussi confirmé par Fig.3.6 et Fig.3.10(a) où il apparaît que le ST-SCS a approximativement le même niveau d'indétectabilité que le stégo-système TCQ. D'un autre côté, Fig.3.10(b) montre qu'au delà d'une certaine valeur du facteur d'étalement τ (approximativement pour $\tau > 8$), l'entropie relative (comme mesure du niveau d'indétectabilité) devient stable et petite, i.e. pour $\tau > 8$, le ST-SCS a un bon niveau d'indétectabilité. Cependant, Fig.3.11 montre que la dérivation de l'entropie relative par rapport à α n'est pas toujours nulle, même si cette dérivé est négative et converge "rapidement" à zéro, donc, l'entropie relative ne prend pas, théoriquement, sa valeur minimale pour toutes les valeurs du paramètre α (voir Fig.3.11(a)) et spécialement pour des valeurs faibles du paramètre α, ce qui correspond au cas d'attaques puissantes de la



(a) (b)

FIGURE 3.11 – (a) La dérivé de l'entropie relative $D(p_S||p_X)$ entre les p.d.f. du cover-signal Gaussien de variance $\sigma_S^2 = 20$ et du stégo-signal en fonction du paramètre α, dans le cas des stégo-systèmes ST-SCS and ST-TCQ pour $\tau = 2$. (b) L'entropie relative avec une puissance d'insertion : $D_1 = 1$ en fonction du paramètre d'étalement τ pour différentes valeurs du paramètre α avec un stégo-système ST-TCQ .

part du gardien actif [1]. De plus, si le paramètre d'étalement $\mathbf{t}$ est public ou alors si le gardien procède à des attaques suivant la direction d'insertion $\mathbf{t}$, la resistance contre les attaques du gardien du ST-SCS stégo-système deviendrait la même que le SCS basique. Donc, nous procédons à des améliorations du stégo-système ST-SCS en remplacant le SCS par le stégo-système TCQ pour obtenir la même indétectabilité que le stégo-schéma TCQ dans le domaine transformé et éliminer tous les effets du paramètre α sur le stégo-schéma.

Lorsque l'indétctabilité des stégo-systèmes est comparée à l'aide de l'évaluation de l'entropie relative entre les p.d.f. du cover-signal et du stégo-signal, on parle de niveau d'indétectabilité et non pas de valeurs précises de l'entropie relative afin d'éviter de parler de systèmes avec des courbes d'entropies relatives très proches telles que la TCQ et le ST-SCS. Ainsi, on se concentre sur les courbes d'entropie relative très éloignées (niveau d'indétectabilité très différent) tel que celles des stégo-systèmes SCS et TCQ.

3.6 Le spread Transform Trellis Coded Quantization

Nous combinons le stégo-système TCQ avec le ST dans le but d'amliorer l'indétectabilité du système, même si le paramètre d'étalement $\mathbf{t}$ est connu, et pour compenser les faiblesses du stégto-système TCQ (une mauvaise résistance contre le gardien actif pour des attaque puissantes) en utilisant le ST. Le système obtenu est appelé le ST-TCQ.

Analyse des performances du ST-TCQ

Théorème3 : Considérons un cover-signal modélisé par la varibale aléatoire S et un signal d'insertion modélisé par la variable aléatoire E, de variance σ_E^2. Lorsque le stégo-système ST-TCQ avec un nombre d'états du treillis très grand est utilisé, la p.d.f. du stégo-signal modélisée par une variable aléatoire X est donnée par la formulation suivante :

$$\begin{aligned} p_X(x) = & \frac{\tau}{2\left(\tau-\alpha\right)} \times \sum_{l,m,t} \int_{-\infty}^{\infty} \int_{0}^{1/2} \\ & \delta\left(\left(l+\frac{m}{2}-\gamma\right)\Delta - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y t-\alpha\left(l+\frac{m}{2}-\gamma\right)\Delta t\right)t+y\right)\right) \\ \times \quad & p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y t-\alpha\left(l+\frac{m}{2}-\gamma\right)\Delta t\right)\right) p_Y(y)\ \mathrm{d}\gamma\ \mathrm{d}y, \end{aligned} \tag{3.46}$$

où :

- S est la variable aléatoire qui modélise le cover-signal et Y est une variable aléatoire auxilaire (l'expression exacte des réalisations y nde la variable aléatoire Y est donnée par Eqn.3.34 dans la preuve du *Théorème 3*),
- α et Δ sont respectivement le paramètre d'optimisation de Costa et le pas du quantificateur scalire utilisé $Q_\Delta(.)$; Ils sont utilisés dans le stégo-système ST-SCS stego-signal avec un paramètre d'étalement τ,
- m est le message à inserer, $\mathbf{t}$ est la direction d'étalement (rappelons que pour rendre les équations plus lisibles nous remplacons $\mathbf{t}$[i] by t) et $l \in \mathbb{Z}$.

Preuve :

Les mêmes modèles et hypothèses concernant le stégo et le cover-signal établis

dans la preuve du *Théorème 1* sont repris dans cette démonstration. Rappelons que le ST-TCQ reviendrait à utiliser le stégo-système TCQ dans le domaine transformé. Donc, nous utiliserons dans ce qui suit la même implémentation du stégo-système TCQ et la même construction du dictionnaire décrit précédement, ainsi, les mots de codes choisis dépendent du message inséré, de l'état du treillis -et plus tard de la direction d'insertion choisie.
En utilisant les résultats obtenus pour le stégo-système ST-SCS, la p.d.f. du signal tatoué avec le ST-TCQ est donnée par :

$$\begin{aligned} p(x|e_t) &= \frac{\tau}{4\cdot(\tau-\alpha)}\cdot \\ &\quad \sum_m \sum_{u_m,t} \\ &\quad \int_{-\infty}^{\infty} \delta\left(u_{e_t,m} - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y\cdot t-\alpha u_{e_t,m}\cdot t\right)\cdot t+y\right)\right) \\ &\cdot \quad p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y\cdot t-\alpha u_{e_t,m}\cdot t\right)\right) p_Y(y)\ \mathrm{d}y. \end{aligned} \tag{3.47}$$

En marginalisant sur les états du treillis possibles E_t, nous obtenons :

$$\begin{aligned} & p_X(x) = \sum_{j=1}^{G} p_X(x|\mathbf{e_t}[j])p(\mathbf{e_t}[j]) \\ = \quad & \frac{\tau}{4\cdot(\tau-\alpha)}\cdot\frac{1}{G}\sum_m \sum_{j=1}^{G} \sum_{u_{\mathbf{e_t}[j],m},t} \\ & \int_{-\infty}^{\infty} \delta\left(u_{\mathbf{e_t}[j],m} - Q_\Delta\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y\cdot t-\alpha u_{\mathbf{e_t}[j],m}\cdot t\right)\cdot t+y\right)\right) \\ \cdot \quad & p_S\left(\frac{\tau}{\tau-\alpha}\left(x+\alpha y\cdot t-\alpha u_{\mathbf{e_t}[j],m}\cdot t\right)\right) p_Y(y)\ \mathrm{d}y, \end{aligned} \tag{3.48}$$

Si nous considérons $\gamma_j = \frac{j}{G/2} \cdot \frac{1}{2}$,

$$
\begin{aligned}
I = & \lim_{G\to\infty} \frac{1}{2} \cdot \frac{1}{G/2} \sum_{j=1}^{G/2} \frac{\tau}{2 \cdot (\tau - \alpha)} \\
\cdot & \sum_{l,m,t} \int_{-\infty}^{\infty} \delta(\Big(l + \frac{m}{2} - \gamma_j \Big) \Delta \\
& - Q_\Delta \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l + \frac{m}{2} - \gamma_j \right) \Delta \cdot t \right) \cdot t + y \right)) \\
\cdot & \; p_S \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l + \frac{m}{2} - \gamma_j \right) \Delta \cdot t \right) \right) \cdot p_Y(y) \; \mathrm{d}y. \qquad (3.49)
\end{aligned}
$$

Puisque cette dernière somme a la forme d'une somme de Riemann, la p.d.f. du stégo-signal peut s'écrire :

$$
\begin{aligned}
& p_X(x) = \frac{\tau}{2 \cdot (\tau - \alpha)} \\
\cdot & \sum_{l,m,t} \int_0^{1/2} \int_{-\infty}^{\infty} \delta(\Big(+ \frac{m}{2} - \gamma \Big) \Delta \\
& - Q_\Delta \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l + \frac{m}{2} - \gamma \right) \Delta \cdot t \right) \cdot t + y \right)) \\
\cdot & \; p_S \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l + \frac{m}{2} - \gamma \right) \Delta \cdot t \right) \right) \cdot p_Y(y) \; \mathrm{d}\gamma \; \mathrm{d}y. \qquad (3.50)
\end{aligned}
$$

Nous remplacons la variable m par ces deux réalisations possibles $\{0, 1\}$ et nous obtenons,

$$
\begin{aligned}
& p_X(x) = \frac{\tau}{2 \cdot (\tau - \alpha)} \\
& \cdot \sum_{l,t} \int_{-1/2}^{1/2} \int_{-\infty}^{\infty} \delta \left((l - \gamma) \Delta - Q_\Delta \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l - \gamma \right) \Delta \cdot t \right) \cdot t + y \right) \right) \\
\cdot & \; p_s \left(\frac{\tau}{\tau - \alpha} \left(x + \alpha y \cdot t - \alpha \left(l - \gamma \right) \Delta \cdot t \right) \right) \cdot p_Y(y) \;\; \mathrm{d}\gamma \; \mathrm{d}y. \qquad (3.51)
\end{aligned}
$$

Fig.3.12 montre que pour différentes valeurs du facteur d'étalement τ et de paramètre de Costa α les p.d.f. obtenues experimentalement suivent ceux du modèle théorique qui sont eux même proches de la p.d.f. du signal hôte.
Lorsque le facteur d'étalement τ devient grand et si le paramètre d'étalement t prend équiprobablement deux valeurs possibles : $\pm 1/\sqrt{\tau}$, Eqn.3.48 de la preuve du

Theorème 3 devient :

$$p_X(x) = \frac{1}{2G} \sum_{j=1}^{G} \sum_{m} \sum_{u_{\mathbf{e_t}[j],m}} \int_{-\infty}^{\infty} \delta\left(u_{\mathbf{e_t}[j],m} - Q_\Delta\left(y\right)\right) \cdot p_S\left(x\right) p_Y(y)\ \mathrm{d}y,$$

où le nombre total des états possibles du treillis est égal à $G \in \mathbb{N}^\star$.

Donc,

$$\begin{aligned} p_X(x) &= \frac{1}{2G} \sum_{j=1}^{G} \sum_{m} \sum_{u_{\mathbf{e_t}[j],m}} \int_{u_{\mathbf{e_t}[j],m}-\Delta/2}^{u_{\mathbf{e_t}[j],m}+\Delta/2} p_S\left(x\right) p_Y(y)\ \mathrm{d}y \;=\; \frac{1}{G} \sum_{j=1}^{G} \int_{-\infty}^{\infty} p_S\left(x\right) p_Y(y)\ \mathrm{d}y \\ &= \int_{-\infty}^{\infty} p_S\left(x\right) p_Y(y)\ \mathrm{d}y \;=\; p_S\left(x\right) \int_{-\infty}^{\infty} p_Y(y)\ \mathrm{d}y \;=\; p_S\left(x\right). \end{aligned} \tag{3.52}$$

Ainsi, la p.d.f. du stégo-signal du ST-TCQ est exactement la même que celle du cover-signal dans le cas limite (le facteur d'étalement τ infini). Cependant, Fig.3.11(b) montre que pour des valeurs du facteur d'étalement supérieures à $\tau = 2$ (valeurs déterminées approximativement de la courbe donnée sur Fif.3.11(b)), l'indétectabilité est bonne et devient faible/stable pour des valeurs du facteur d'étalement supérieures à $\tau = 20$. Aussi, nous obtenons une p.d.f. du stégo-signal qui soit proche de celle du cover-signal pour des valeur du facteur d'étalement τ modérées. Tel qu'il est montré dans Fig.3.7, pour la même attaque du gardien actif (suivre les deux Rectangles de Fig.3.7 : ligne continue pour le stégo-système TCQ et la ligne discontinue pour le ségo-système ST-TCQ), même si l'insertion avec le stégo-système TCQ permettrait un gain en capacité, il ne resiste aux attaques du warden (il y a approximativement 2×10^5 de bits erreurs pour 10^6 bits transmis) alors que le stégo-système ST-TCQ, avec un facteur d'étalement τ égal à 10, résiste parfaitement aux attaques du gardien, d'ailleurs, il y a seulement deux bits erronés pour 10^6 transmis, de plus, le niveau d'indétectabilité est meilleur pour le stégo-système ST-TCQ que pour le stégo-système TCQ dans ce cas.

De Fig.3.7, on peut noter que les perfomances globales du ST-TCQ sont du même niveau que celles du stégo-système ST-SCS. Cependant, le gain obtenu avec le ST-TCQ en terme d'indétectabilité est indiscutable, puisque l'indétectabilité dans le domaine transformé est la ùêùe que celle du stégo-système TCQ (ST-SCS induit distord le stégo-signal dans le domaine tranformé) et, contrairement au stégo-système ST-SCS, il n'est pas affecté par le paramètre α (voir Fig.3.11 et Fig.3.10).

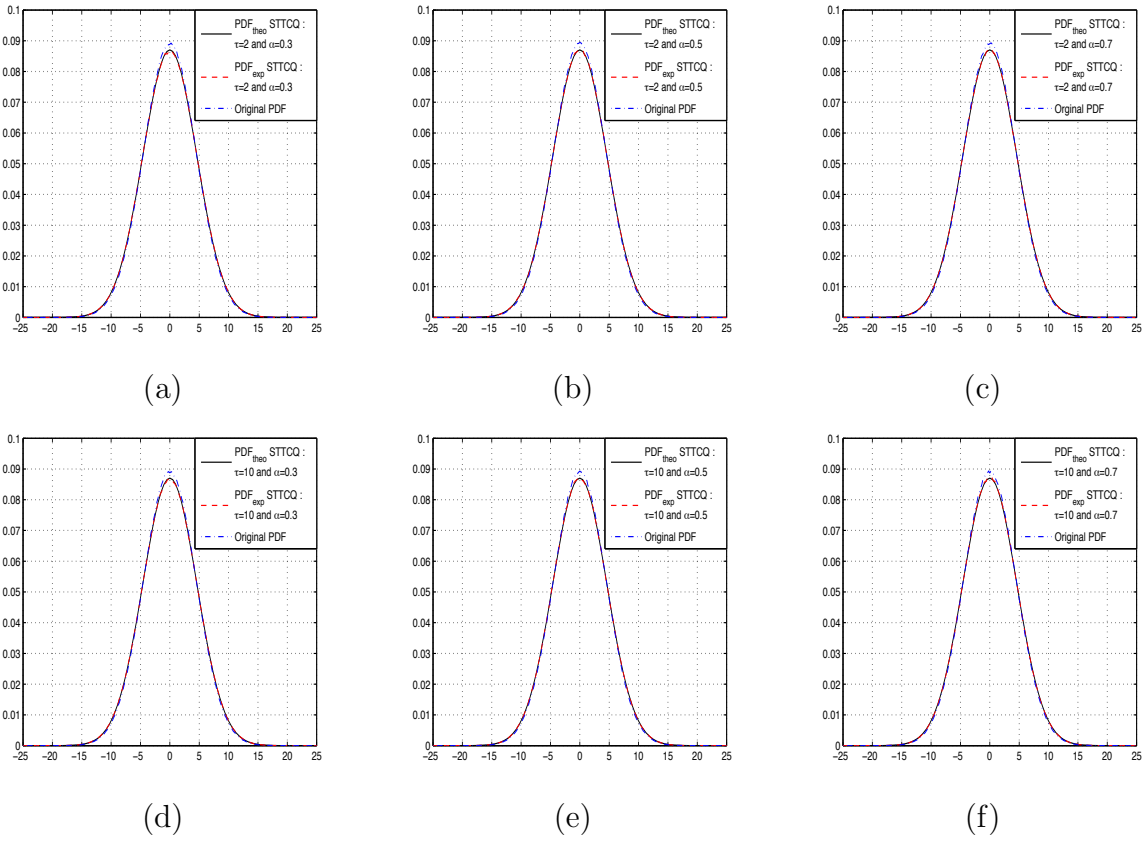


FIGURE 3.12 – Fonctions densités de probabilité d'un cover-signal Gaussien, de variance $\sigma_S^2 = 20$, et du stego-signal en utilisant le stégo-schéma ST-TCQ, avec une puissance d'insertion $D_1 = 1$, pour $\tau = 2$ avec des différentes valeurs du paramètre α : (a) $\alpha = 0.3$, (b) $\alpha = 0.5$ and (c) $\alpha = 0.7$; et pour $\tau = 10$ avec (d) $\alpha = 0.3$, (e) $\alpha = 0.5$ et (f) $\alpha = 0.7$.

Notons que sur Fig.3.6(a), le stégo-système ST-SCS apparait comme étant mieux que le ST-TCQ en termes d'indétectabilité. Ceci ne permet pas de conclur que le ST-SCS stégo-message est moins détectable que le ST-SCS , puisque la petite différence de l'entropie relative est due aux conditions expérimentales spécifiques : cover-signal Gaussien avec contraintes stéganographiques.

Même si la p.d.f. simple $p_X(x)$ n'est pas distordue, ceci ne signifie absolument pas que la condition d'indétectabilité est complètement vérifiée. En fait, la dépendance entre pixels peut être "perturbée" par l'insertion du stégo-message. Dans le but d'evaluer les effets de l'insertion du stégo-message sur la corrélation existante

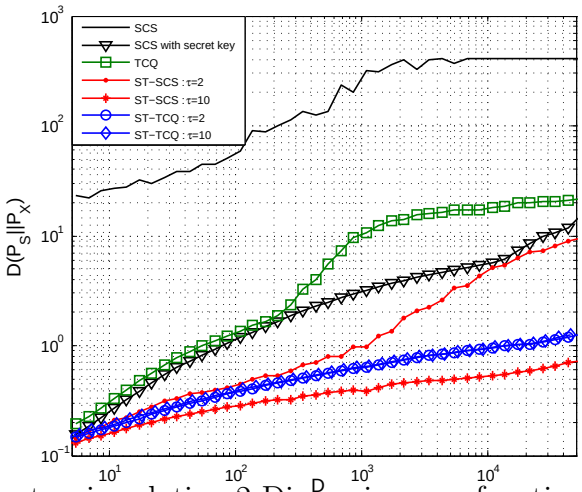
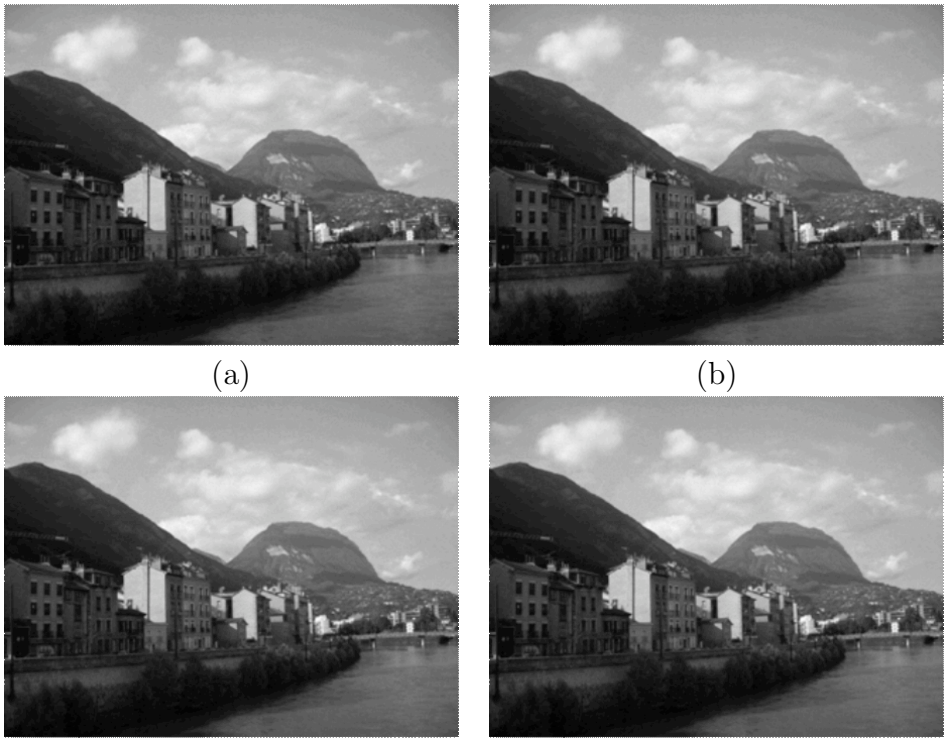


FIGURE 3.13 – L'entropie relative 2-Dimensions en fonction de la puissance d'insertion D_1, dans le cas de stégo-systèmes SCS, TCQ, ST-SCS and ST-TCQ ; nous utilisons 100 images réelles de taille 350×350 pixels.

entre les pixels de la stégo-image, nous calculons l'entropie relative entre la densité de probabilité conjointe (entropie relative 2-Dimensions) des pixels de la stégo-image et de la cover-image (En fait, nous utilisons les estimations des p.d.f. puisque nous travaillons avec des images dont le nombre de pixels est limité). Nous considérons le même contexte que dans le cas du calcul de l'entropie relative 1-Dimension expérimentale.
Les résultats du calcul de l'entropie relative 2-Dimension pour les stégo-systèmes : SCS, TCQ, ST-SCS et ST-TCQ sont données sur Fig.3.13. On constate que pour les images réelles, le stégo-système SCS affecte la densité de probabilité conjointe du stégo-signal plus que les autres systèmes même pour des puissances d'insertion relativement faibles. Notons qu'il est très important d'utiliser des images réelles pour ces dernières expériences pour mettre en évidence l'effetdu tatouage sur la corrélation entre les pixels des images.

Sur Fig.3.14, nous remarquons que la fidélité des stégo-images tatouées avec les stégo-systèmes : SCS, TCQ, ST-SCS et ST-TCQ sont approximativement du même niveau -puisque le PSNR est aux alentours de 40dB- lorsque la puissance d'insertion n'est pas importante. D'un autre côté, Tab.3.1 présente une comparaison entre l'entropie relative obtenue et le taux d'erreur binaire lorsqu'on utilise une cover-image de taille 320×240 pixels, tel que le rapport puissance du cover-signal



(a) (b)

(c) (d)

FIGURE 3.14 – Une stégo-image (Grenoble) de taille 320×240 pixels avec les stégo-systèmes : (a) SCS, (b) TCQ, (c) ST-SCS for $\tau = 10$ and (c) ST-TCQ for $\tau = 10$, tel que le ratio entre les puissance du signal hôte et celle de l'insertion sont égales à 35 dB.

	SCS	TCQ	ST-SCS	ST-TCQ
Payload/Taille d'image (i.e. bits/pixel)	1	1	0.1	0.1
b.e.r.	0.2278	0.4194	0.0039	$< 10^{-5}$
$D(p_S \| \| p_X)$	0.9910	0.0094	0.0093	0.0072

TABLE 3.1 – Taux d'erreurs binaires b.e.r. (Bit error rate) et l'entropie relative pour les stégo-systèmes SCS, TCQ, ST-SCS ($\tau = 10$) et ST-TCQ ($\tau = 10$) où le message est inséré dans des images réelles de taille 320×240 (Fig. 3.14). Le rapport puissance d'insertion sur la puissance du gardien est égal à 0 dB et le rapport de la puissance du cover-signal sur puissance d'insertion est égal à 35 dB.

sur puissance d'insertion est égal à 35 dB, aussi, le rapport puissance d'insertion sur puissance d'attaque est égal à 0dB.

3.7 Conclusions

Les contributions majeures que nous avons présentées dans ce chapitre peuvent être résumées dans les points suivants :

- une analyse complète dans le contexte du gardien actif a été effectuée pour différents systèmes basés sur les travaux de Costa proposés pour le codage du canal,
- grâce à l'entropie relative 2-Dimensions, les effets de quatre différentes techniques d'insertion (SCS, TCQ, ST-SCS et ST-TCQ) sur la corrélation entre pixels d'une image sont étudiés et mis à l'evidence,
- un nouveau stégo-schéma appelé ST-TCQ résultant de la combinaison du spread transform et la TCQ a été proposé,
- les expressions analytiques des fonctions densités de probabilité ont été développées pour différents stégo-systèmes : SCS, TCQ, ST-SCS et ST-TCQ.

Dans le contexte de la stéganographie avec gardien actif, nous avons montré les limites et les avantages de plusieurs stégo-schémas avec information adjacente en termes d'indétectabilité, de résistance face aux attaques du gardien actif et la capacité stéganographique. Pour chaque système , des résultats expérimentaux ont été utilisés pour valider les modèles théoriques. Pour le SCS, le stégo-signal est régulièrement partitionné et plusieurs aréfacts dans la p.d.f. du stégo-signal apparaissent. Nous avons prouvé théoriquement plusieurs points expliqués souvent d'une manière ″intuitive″. A cause de ces observations, nous avons proposé une analyse de deux autres stégo-systèmes. Le premier basé sur un partitionnement pseudo-aléatoire (un système basé sur la TCQ), qui permet d'obtenir un stégo-système plus fléxible et avec une meilleure indétectabilité. Le second système est basé sur la combinaison du SCS avec le spread transform (ST-SCS), nous avons prouvé le bon niveau d'indétectabilité et la très bonne résistance face aux attaques du gardien, aussi, nous avons montré le très bon compromis Capacité-Résistance-Indétectabilité offert par ce dernier. Ayant trouvé que les perfomances du ST-SCS sont les mêmes que le SCS dans le domaine transformé, nous avons proposé donc une nouvelle combinaison qui est celle du ST avec la TCQ. Le ST-TCQ offre de très bonnes perfomances proches

de celles du ST-SCS avec un meilleur niveau d'indétectabilité. D'un autre côté, le ST-TCQ permet en plus une très bonne indétectabilité égale à celle de la TCQ dans le domaine tranformé.

Chapitre 4

Tatouage numérique

4.1 Introduction

Dans ce chapitre, nous traitons principalement de la sécurité des systèmes de tatouage numérique. Après une présentation des caractéristiques génériques d'un système de tatouage numérique, nous étudions la sécurité des systèmes informés basés sur la quantification. Nous nous intéressons d'abord aux attaques par estimation, où nous présentons les résultats théoriques que nous avons obtenus pour l'évaluation de la sécurité du tatouage dans le cas particulier des systèmes informés. Ensuite, nous présentons une étude de la sécurité de quelques systèmes informés et l'effet de l'ajout d'une couche ST. Dans la deuxième partie, nous étudions les attaques par effacement de la marque. Nous nous intéressons particulièrement à l'attaque TFA (Temporal Frame Averaging). Nous présentons les résultats théoriques développés pour évaluer l'effet de l'attaque par moyennage sur le ST, puis, nous comparons ces résultats à ceux obtenus des simulations sur des signaux réels (vidéo). Enfin, nous introduisons une nouvelle notion qui est *"la cicatrice du tatouage numérique"* que nous présentons comme une solution pour contrer les attaques par élimination (effacement) de la marque. Notons que certaines parties de ce chapitre ont été publiées dans les les articles suivants : [47] [48] [49]

4.2 Généralités sur le tatouage numérique robuste

Le tatouage numérique est une discipline très récente. Sa naissance remonte au début des années 90 avec l'article de Tanaka [50]. Digital watermarking (tatouage

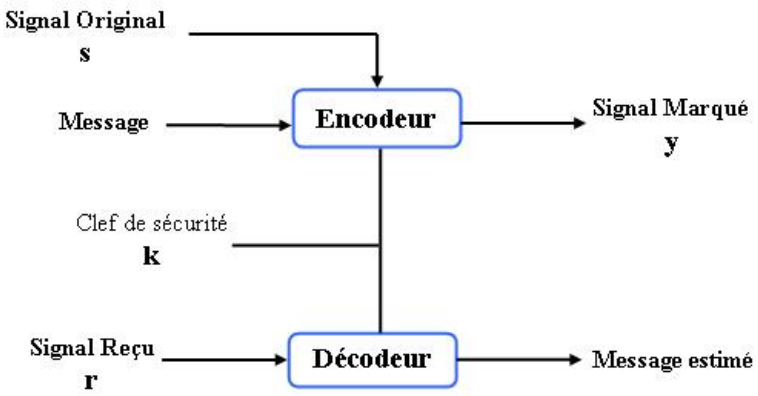


FIGURE 4.1 – Schéma général d'un système de tatouage numérique.

numérique) fut pour la première fois employé en 1993 par Tirkel [51] (le terme utilisé est originaire du Japon "denshi Sukashi" qui se traduit "Elctronic watermark").

4.2.1 Définition

Le tatouage numérique consiste à transmettre un message m via un support hôte s. Ainsi, le signal est modifié de façon à permettre au décodeur d'extraire le message sans que les caractéristiques du signal hôte soient modifiés, autrement dit, l'insertion de l'information doit être parfaitement imperceptible.
L'opération de tatouage peut être résumée dans le schéma de Fig.4.1. La marque est choisie dans l'alphabet $\mathcal{M}$, i.e., $m \in \{1, ..., N\}$, où N est un entier naturel. Le signal tatoué : $x = s + w_m$, où w_m est le signal watermark obtenu à partir du message m, est transmit sur le canal où il subit diverses distorsions et dégradations. Ensuite, le décodeur donne une estimation du message à partir du signal reçu.

4.2.2 Caractéristiques du tatouage numérique

Tout schéma de tatouage découle inévitablement d'un compromis entre les principales caractéristiques du tatouage : Robustesse, sécurité, capacité et transparence. Ce compromis répond à certaines conditions dictées par un ensemble d'éléments (le niveau de menace qui pèse sur notre marque, l'environnement dans lequel évolue le document marqué,etc...) qui sont liés à l'application à laquelle est destiné le système.

Dans la suite de ce chapitre, nous présentons brièvement la sécurité d'un système de tatouage, ainsi que le compromis entre la robustesse, la transparence et la capacité dans le cas de systèmes informés basés sur la quantification. Ces trois propriétés ont largement été étudiées dans la bibliographie pour les systèmes informés [1] [15] [46] [52] [5], leurs définitions sont données dans le deuxième chapitre qui présente l'état de l'art. Cependant, nous détaillerons la partie sécurité qui est la moins étudiée, en particulier, dans le cas des systèmes informés et qui est le point qui est propre au contexte du tatouage numérique.

Sécurité

On définit la sécurité d'un système de tatouage comme étant la difficulté qu'aura un utilisateur mal-intentionné à retrouver l'information insérée ou à supprimer la marque et récupérer le signal original. La sécurité du watermarking repose, généralement, sur une ou plusieurs clefs cryptographiques qu'on utilise à l'insertion et à l'extraction de notre information. Par exemple, dans certains schémas de tatouage, l'information est insérée sous forme d'un signal produit par un générateur pseudo-aléatoires dont la combinaison d'entrée représente la clef secrète. Il existe deux niveaux de sécurité, dans le premier, un utilisateur non-autorisé ne pourra pas détecter, décoder ou lire le message inséré. Dans le second, le message inséré est crypté et ne peut pas être lu qu'à l'aide d'une clef tenue secrète. Autrement dit, il faut que le système de tatouage résiste à toute attaque venant d'une personne mal-intentionnée et la clef de cryptage doit rester une défense ultime seulement !.
Lorsqu'on parle de sécurité, cela revient à parler des attaques qui mettent en péril la sécurité de la marque insérée. Ces attaques peuvent être classées en deux grandes catégories :

1. **Les attaques par estimation :** parfois appelées *attaques informées* [53], ce sont celles où l'attaquant procède d'abord à une estimation d'un ou plusieurs éléments du système de tatouage numérique, puis, en se basant sur cette information il élimine la protection du tatouage numérique.
2. **Les attaques par élimination de la marque :** elles regroupent toutes les attaques dont le but est d'éliminer la protection sans altérer la qualité perceptuelle du média attaqué. Généralement, elles s'effectuent de manière aveugle où l'attaquant n'utilise pas forcément les particularités du système de tatouage.

4.2.3 Compromis Robustesse-Capacité-Imperceptibilité : Cas des systèmes informés

Nous avons pris comme référence le système SCS [1] dans ce chapitre parce que ce système est basique par rapport aux autres systèmes basés sur la quantification. De plus, il est très perfomant (voir [1] ou [5]), surtout, en termes de capacité et de robustesses, par rapport, aux systèmes non-informés, en particulier, le fameux spread spectrum watermarking [26], comme expliqué dans le chapitre sur l'état de l'art.
Fig.4.2 montre les performances des systèmes informés décrits précédement et leurs compromis Robustesse-Fidélité-Capacité. On peut voir que sur les trois propriétés Capacité-Fidélité-Robustesse le SCS est la TCQ sont à peu près au même niveau, même si nous constatons un avantage de la TCQ sur le SCS pour des puissances de distorsion faibles et inversement pour des puissances plus importantes. D'un autre côté, le ST-SCS et le ST-TCQ ont aussi des performances très proches. En plus, nous constatons qu'ils ont une meilleure robustesse (pour une puissance d'insertion fixe) et une très bonne fidélité (pour une puissance de bruit fixe) qui s'améliorent lorsque le facteur d'étalement τ est grand. Cependant, la capacité diminue en fonction du facteur d'étalement utilisé. Au regard des 3 propriétés Capacité-Imperceptibilité-Robustesse, le système ST-TCQ est celui qu'offre le meilleur compromis par rapport aux systèmes de watermarking SCS, TCQ, ST-SCS.
Dans la suite nous étudions la dernière propriété d'un système de watermarking qui est la sécurité du tatouage numérique. Comme cette dernière dépend fortement des attaques subies par la marque, l'étude est divisée en deux parties : Les attaques par estimation et les attaques par élimination de la marque.

4.3 Attaques par estimation de la marque

Ce sont les attaques qui sont considérées le plus souvent pour l'évaluation de la sécurité, comme il a été proposé dans l'article de Cayre et al. [54]. Elles se décomposent généralement en deux étapes, la première est une étape d'apprentissage dans laquelle l'attaquant estime la clef secrète du système de tatouage. La deuxième étape consiste à utiliser le secret pour casser la sécurité du tatouage. Dans le cas du copyright par exemple, l'attaquant estime la clef de cryptage en analysant plusieurs copies marquées, puis, il utilise cette clef pour insérer un autre copyright afin de

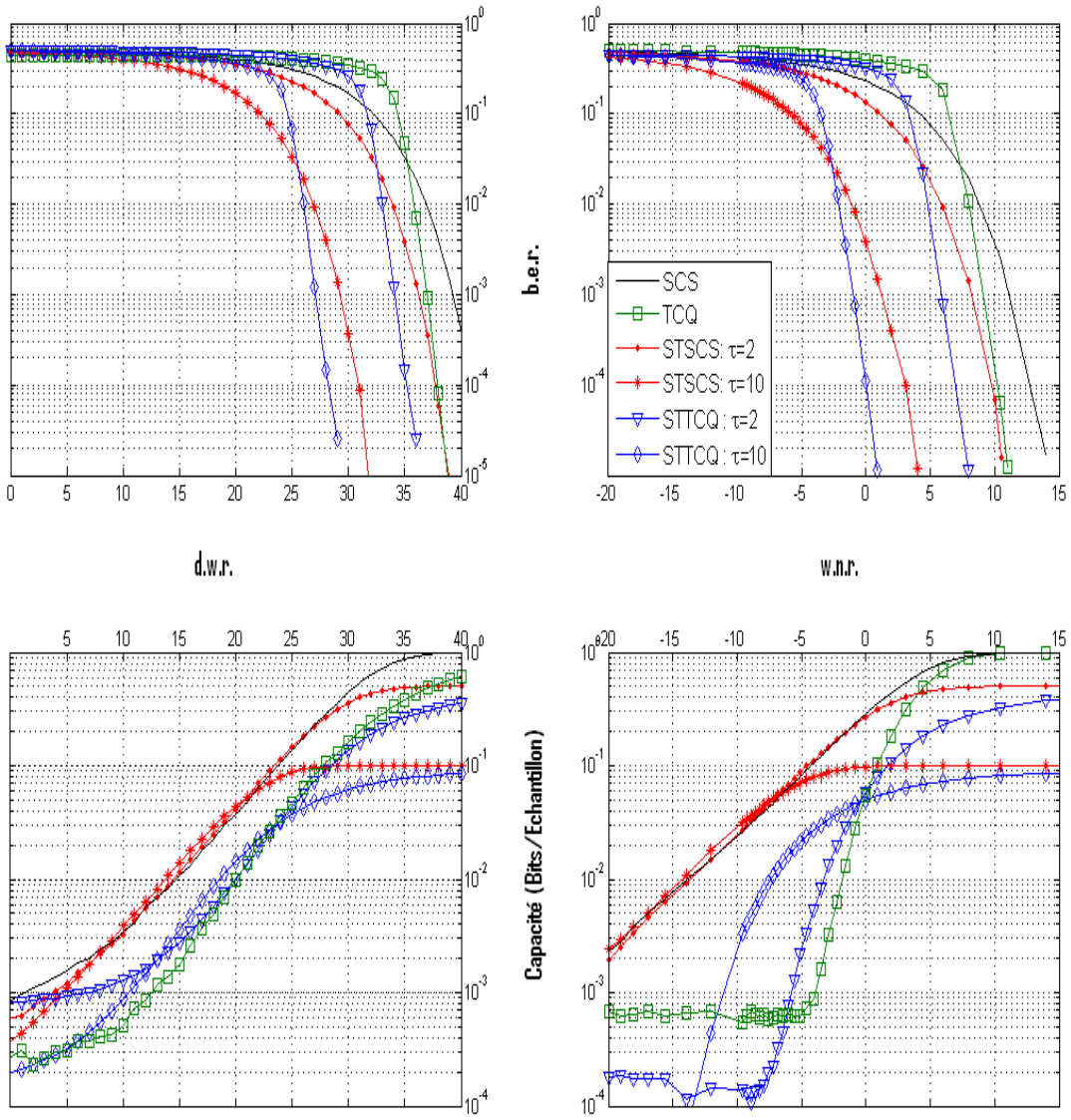


FIGURE 4.2 – Perfomances des systèmes : SCS, TCQ, ST-SCS et ST-TCQ données par les courbes de variation Imperceptibilité-Capacité-Robustesse.

rendre impossible l'identification du véritable propriétaire.

4.3.1 Notions de cryptographie utilisées dans le tatouage numérique

Souvent la cryptographie et le watermarking sont vus comme deux domaines disjoints et parfois même ″concurrents″. En réalité, ils sont très complémentaires, puisqu'ils ont le même objectif qui est d'assurer la sécurité des échanges de documents. Dans la suite, nous présentons deux principes tirés de la cryptographie qui sont très importants pour l'analyse de la sécurité des systèmes de watermarking :

Principe de Kerckhoffs

Le principe a été formulé comme suit : ≪Toute méthode de chiffrement est connue de l'ennemi, la sécurité du système ne dépend que du choix des clefs ≫. Le principe de Kerckhoffs a été formulé dans les numéros de janvier et février 1883 du ″journal des sciences militaires″ [55]. Le professeur Auguste KERCKHOFFS traitait du problème de communication entre les différents chefs d'armée. Il a démontré qu'il est impossible de garder un algorithme de cryptage secret indéfiniment. Ainsi, il a pu établir que pour qu'un crypto-système soit sécure, il faudrait que tout ce qui concerne ce système doit être public, sauf, la clef de cryptage qui reste secrète. Ceci est le contraire de ce qui est appelé *la sécurité par l'obscurité* [56] qui consiste à garantir la sécurité des systèmes en gardant les algorithmes de cryptage secret.
En watermarking, la situation est très similaire à celle de la cryptographie. Les algorithmes et les paramètres utilisés dans les systèmes de tatouage sont publics, uniquement, la clef secrète est cachée pour sécuriser la marque. Généralement, cette clef est une suite numérique obtenue à l'aide d'un générateur pseudo-aléatoire, ou alors, d'un paramètre important du système de tatouage dont la connaissance est nécessaire pour le décodage de l'information insérée.

Principe de Shannon

Dans ce cas, il est supposé qu'une clef secrète est utilisée pour protéger plusieurs documents, ainsi, l'attaquant observe plusieurs contenus protégés par cette seule clef cachée. Un système est dit *parfaitement secure*, si et seulement si, aucune fuite d'information, liée à la clef de sécurité, ne pourrait se produire des documents protégés.

Sinon, *le niveau de sécurité* est défini comme le nombre d'observations nécessaires à l'estimation de la clef secrète utilisée dans le système de sécurité. Ainsi, plus la fuite d'information depuis les observations est réduite plus le niveau de sécurité du système est élevé.
Cayre et al. [54] ont proposé de transposer cette approche pour la sécurité du watermarking, afin d'évaluer la sécurité d'un système de tatouage, où une clef secrète est utilisée lors de l'insertion du tatouage numérique. Il est supposé que l'attaquant a en sa possession plusieurs copies marquées avec différents tatouages mais où une seule clef secrète est utilisée. Ainsi, un système de tatouage est *parfaitement sécure* s'il n' y avait aucune fuite d'information des copies marquées. Par analogie à la cryptographie, le niveau de sécurité d'un système de tatouage est donné par le nombre d'observations nécessaires à l'estimation de la clef secrète utilisée lors de l'insertion du tatouage.

4.3.2 Mesures de la sécurité d'un système de tatouage numérique

Comme dans l'article [54], la sécurité du tatouage numérique est calculée par rapport à la résistance contre les attaques par estimation. Ainsi, la clef de sécurité doit rester secrète : aucune personne autre que celle autorisée ne peut connaître le secret du tatouage.
Dans l'article [54], les auteurs ont établi que la sécurité du tatouage numérique est donnée par la taille des observations disponibles au niveau de l'attaquant. Les auteurs ont donné des outils théoriques tirés de la théorie de l'information dans le but d'évaluer le niveau de sécurité d'un système de watermarking. Ainsi, nous présentons les deux mesures de sécurité possibles dans le cas d'une attaque par estimation :

Mesure de Shannon

Dans le cas où la clef secrète K est une variable discrète, ou plus usuellement un mot binaire, l'entropie $H(\mathbf{K})$ mesure l'incertitude des attaquants sur la vraie valeur de la clef secrète utilisée $\mathbf{k}$. Lorsque l'attaquant observe $\mathbf{O}^{N_0}$, l'incertitude sur la clef secrète est évaluée avec l'entropie conditionnelle que Shannon nomme l'*équivocation*, elle est donnée par la formule suivante :

$$H((\mathbf{K}|\mathbf{O}^{N_0})) = H(\mathbf{K}) - I(\mathbf{K};\mathbf{O}^{N_0}). \tag{4.1}$$

La fuite d'informations est mesurée par l'information mutuelle entre les observations et la clef secrète. Plus la fuite d'information est importante moins sera l'incertitude de l'attaquant sur la clef secrète.
L'équivocation est décroissante en fonction de N_0 : elle varie entre l'entropie $H(\mathbf{K})$ à 0. Une équivocation nulle veut dire que l'ensemble des clefs secrètes possibles utilisées pour le tatouage est réduit à un ensemble d'un unique élément. Ainsi, il est possible de mesurer le niveau de sécurité $N_0^\star$ du système de tatouage utilisé.
Malheureusement, ces outils d'évaluation de la sécurité ne peuvent être utilisés sur tous les systèmes de tatouage numérique. Dans la littérature [43] [57], il est connu que l'entropie (ou l'entropie conditionnelle) d'une variable aléatoire continue ne mesure pas une quantité d'information. L'information mutuelle $I(\mathbf{K};\mathbf{O}^{N_0})$ est toujours une mesure pertinente de la fuite d'information, mais l'interprétation physique de l'équivocation n'est plus l'incertitude lorsque la clef secrète est une variable continue, puisque l'équivocation peut prendre une valeur positive ou négative ce qui annule le concept d'unicité de la distance.

Mesure de Fisher

Cette mesure a été proposée pour l'évaluation de la sécurité là où la mesure de Shannon ne peut le faire, c.à.d., le cas où le système de tatouage utilise une clef secrète continue.
En statistiques, Fisher est considéré comme l'un des pionniers dans la mesure de la quantité d'information en s'appuyant sur les observations d'un paramètre inconnu à estimer. Si on suppose que les observations représentées par le vecteur de variable aléatoire $\mathbf{O}$ ayant une distribution de probabilité qui dépend du vecteur paramètres $\boldsymbol{\theta}$. *La matrice de l'information de Fisher* : FIM (Fisher Matrix Information) du paramètre $\boldsymbol{\theta}$ est définie comme suit

$$FIM(\boldsymbol{\theta}) = \mathbb{E}\psi\psi^T \qquad avec \qquad \psi = \nabla_{\boldsymbol{\theta}} \log p_{\mathbf{O}}(\mathbf{o};\boldsymbol{\theta}), \tag{4.2}$$

où $\mathbb{E}[.]$ représente l'opérateur espérance mathématique d'une variable aléatoire et $\nabla_{\boldsymbol{\theta}}$ est l'opérateur gradient vectoriel défini par $\nabla_{\boldsymbol{\theta}} = (\partial/\partial\boldsymbol{\theta}[1], ..., \partial/\partial\boldsymbol{\theta}[N_\theta])^T$, tel que N_θ représente la taille du vecteur paramètre $\boldsymbol{\theta}$. Le théorème de Cramer-Rao donne une borne inférieur $\mathcal{R}_{\boldsymbol{\theta}}$ de la matrice de covariance d'un estimateur non-biaisé du vecteur paramètre $\boldsymbol{\theta}$, tant que la FIM reste inversible,

$$\mathcal{R}_{\boldsymbol{\theta}} \geq FIM(\boldsymbol{\theta})^{-1}. \tag{4.3}$$

Dans notre contexte, le vecteur des paramètres à estimer peut être le watermark inséré ou alors la clef secrète utilisée lors de l'insertion de la marque (voir Fig.4.1). L'interprétation physique de Eqn.4.3 est : plus la fuite d'information du paramètre secret est importante mieux sera la précision de l'estimation de ce paramètre.
La FIM est une mesure additive de l'information lorsque les observations disponibles sont statistiquement indépendantes. Si $\hat{\boldsymbol{\theta}}$ représente l'estimation du vecteur paramètre $\boldsymbol{\theta}$, alors, l'erreur quadratique moyenne $\mathbb{E}\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\}$ serait la trace de $\mathcal{R}_{\boldsymbol{\theta}}$, donnée par Eqn.4.3, qui décroît lorsque le nombre d'observation N_0 croît. L'estimation est d'autant plus précise lorsque les observations sont indépendantes et en nombre important. D'un autre côté, plus le nombre d'observations nécessaires à une estimation précise du paramètre $\boldsymbol{\theta}$ est important, plus la difficulté du côté attaquant sera importante. Ainsi, si la clef secrète est le paramètre à estimer alors le niveau de sécurité serait donné par $N_0^\star$ correspondant au nombre de copies nécessaires à une estimation précise de la clef secrète utilisée lors de l'insertion du tatouage numérique.

4.3.3 Sécurité au sens de Shannon d'un système de tatouage lorsque les observations sont des copies marquées indépendantes

Lorsque la clef secrète est une variable discrète, l'information mutuelle entre les N_0 observations et le paramètre secret -à estimer par l'attaquant- est donnée par :

$$I(X^{N_0}; K) = \sum_k \int_{x_1} \int_{x_2} ... \int_{x_{N_0}} p(x_1, x_2, ..., x_{N_0}, k) \log(\frac{p(x_1, x_2, ..., x_{N_0}, k)}{p(x_1, x_2, ..., x_{N_0})p(k)}), \tag{4.4}$$

où K est la variable aléatoire qui modélise la clef secrète et $k[i] \in [-1/2, 1/2]$, $i = 1, ..., N_0$ comme dans l'article [2].
Puisque les variables aléatoires X_i, $i = 1, ..., N_0$ sont indépendantes, identiquement distribuées (i.i.d.), tel que x_i modélise la i^{me} copie (i^{me} observation), sa fonction densité de probabilité conjointe peut s'écrire comme suit

$$p(x_1, x_2, ..., x_{N_0}) = p(x_1) \cdot p(x_2) ... \cdot p(x_{N_0}).$$

D'un autre côté, un échantillon marqué avec une clef secrète fixée est formulé comme

$$x_i = s_i + w_{i,k}, i = 1, ..., N_0, \tag{4.5}$$

tel que $w_{i,k}$ est le tatouage de la i^{me} observation pour une seule clef secrète k. Notons que pour clef secrète fixée k, les échantillons marqués x_i, $i = 1, ..., N_0$ restent indépendants, alors

$$p(x_1, x_2, ..., x_{N_0}|k) = p(x_1|k) \cdot p(x_2|k) ... \cdot p(x_1|k), \tag{4.6}$$

par marginalisation, nous obtenons,

$$\begin{aligned} I(X^{N_0}; K) = & \underbrace{\sum_k \int_{x_1} p(x_1, k) \log \left[\frac{p(x_1|k)}{p(x_1)} \right]}_{I(X_1;K)} + \underbrace{\sum_k \int_{x_2} p(x_2, k) \log \left[\frac{p(x_2|k)}{p(x_2)} \right]}_{I(X_2;K)} \\ + \quad & ... + \underbrace{\sum_k \int_{x_{N_0}} p(x_{N_0}, k) \log \left[\frac{p(x_{N_0}|k)}{p(x_{N_0})} \right]}_{I(X_{N_0};K)}. \end{aligned} \tag{4.7}$$

Ainsi, la fuite d'information des N_0 observations, dans le cas des schémas de tatouage informés, est la somme des fuites d'informations de chaque observation indépendantes, i.e.,

$$I(X^{N_0}; K) = \sum_{i=1}^{N_0} I(X_i; K). \tag{4.8}$$

Lorsque l'équivocation est égale à zéro, ceci veut dire qu'il n'y a plus de secret. Si les mêmes schémas de tatouage informés sont utilisés pour toutes les observations et en considérant Eqn.4.1, la quantité d'observations nécessaires pour estimer complètement la clef secrète devient,

$$N_0 = \frac{H(K)}{I(X; K)}. \tag{4.9}$$

Dans la suite de cette partie, nous utilisons Eqn.4.9 pour mesurer le niveau de sécurité des systèmes de tatouage considérés.

Notons que jusqu'à présent, il n'existe pas de travaux sur l'évaluation de la

sécurité du système QIM, et plus généralement, le tatouage SCS avec clef secrète continue. Ceci est dû à la continuité de la clef secrète utilisée dans ces systèmes. De plus, il est impossible de mesurer l'information de Fisher pour ce cas à cause de la non-dérivabilité des densités de probabilité des ces systèmes.
Dans le paragraphe suivant, nous donnerons une analyse détaillée de la sécurité du tatouage QIM et QIM, puis, nous présentons une version sécurisée du système de tatouage basé sur le Treillis Coded Quantization (TCQ).

4.3.4 Sécurité des systèmes basés sur la quantification

Le système de tatouage QIM : Quantization Index Modulation

Le système de tatouage QIM [15] est basé sur la quantification scalaire du signal hôte. C'est un cas particulier du schéma SCS [1], parfois appelé DC-QIM. Considérons un message binaire $\mathbf{m}$ à insérer et un signal hôte $\mathbf{s}$. Dans le cas du système de tatouage QIM, deux dictionnaires sont définis en utilisant deux quantificateurs scalaires décalés :

$$\mathcal{U}_0[i] = \{n\Delta + \mathbf{k}[i]\Delta,\, n \in \mathcal{Z}\} \quad \text{et} \quad \mathcal{U}_1[i] = \{n\Delta + \mathbf{k}[i]\Delta + \frac{\Delta}{2},\, n \in \mathcal{Z}\}, \tag{4.10}$$

où Δ est le pas de quantification, $\mathbf{k}[i]$ représente la i^{me} composante de la clef secrète continue -donnée par le vecteur $\mathbf{k}$-, tel que $\mathbf{k}[i] \in [-1/2, 1/2]$, et $\mathcal{Z}$ est l'ensemble des entiers.
Selon le bit message $\mathbf{m}[i]$ que l'on voudrait insérer, l'un des deux dictionnaires, donnés par Eqn.4.10, est choisi, ainsi, le mot de code $\mathbf{u}^\star[i]$ le plus proche de l'échantillon hôte $\mathbf{s}[i]$ est sélectionné. Évidemment, l'échantillon marqué sera donné par : $\mathbf{x}[i] = \mathbf{u}^\star[i]$.
Nous utilisons la formule d'insertion de la marque donnée dans [1] et fixons le paramètre de Costa α à 1. Cette de valeur de α correspond à un tatouage où le canal de transmission est sans bruit (voir les articles [1] et [15] pour plus de détails sur la dépendance entre le paramètre α et la puissance du bruit). Pour extraire le message inséré, le décodeur calcule un critère de décision noté $\mathbf{y}$, ce critère est formulé avec l'expression suivante et qui a été donnée dans l'article [1] :

$$\mathbf{y} = Q_\Delta(\mathbf{x} - \mathbf{k}\Delta) - (\mathbf{x} - \mathbf{k}\Delta), \tag{4.11}$$

où $\mathbf{x}$ représente le signal reçu sans distorsion, due au canal de transmission par exemple, et $Q_\Delta(.)$ est le quantificateur scalaire de pas Δ utilisé lors du tatouage.
Un simple développement montre que $\mathbf{m}[i] = 1$ lorsque $\|\mathbf{y}[i]\| = \frac{\Delta}{2}$ et $\mathbf{m}[i] = 0$ lorsque $\|\mathbf{y}[i]\| = 0$. Cependant, si le canal est bruité, le critère de décision est utilisé comme suit :

$$m = \begin{cases} 1 \text{ if } |\mathbf{y}[i]| > \Delta/4 \\ 0 \text{ if } |\mathbf{y}[i]| \leq \Delta/4 \end{cases} . \tag{4.12}$$

Pour plus de détails concernant le critère de décision et les statistiques découlant des conditions des systèmes de tatouage, voir l'article *Data-Hiding Codes* [16] de Pierre Moulin.

Est-il possible que l'attaquant contourne la sécurité obtenue à l'aide de la clef continue ?

La réponse est : OUI.
Malheureusement, l'attaquant est supposé détenir une copie tatouée non bruitée, de plus, il est supposé qu'il ait accès à tous les paramètres d'insertion à l'exception de la clef secrète (principe de Kerckhoffs [55]). Donc, il serait possible pour l'attaquant de considérer que la clef secrète est un bruit ajouté à la copie de tatoué. Ceci est d'autant plus vrai lorsque le tatouage est additif ou alors qu'il peut être représenté par un ajout d'un signal (représentant la marque) à l'original , comme c'est le cas pour les systèmes à information adjacente. Si l'attaquant voudrait estimer le message sans connaitre exactement la clef secrète, il est possible pour lui de calculer le critère de décision $\mathbf{y}[i]$ pour le i^{me} échantillon marqué d'une manière $''$aveugle$''$ (le critère de décision est calculé comme si aucune clef secrète ne fut utilisée), ainsi, $\mathbf{y}$ est calculé comme suit,

$$\mathbf{y}[i] = Q_\Delta(\mathbf{x}[i]) - \mathbf{x}[i] = Q_\Delta(n\Delta + \mathbf{k}[i]\Delta + \mathbf{m}[i]\frac{\Delta}{2}) - (n\Delta + \mathbf{k}[i]\Delta + \mathbf{m}[i]\frac{\Delta}{2}), \quad n \in \mathcal{Z}, \tag{4.13}$$

sachant que : $\mathbf{x}[i] = n\Delta + \mathbf{k}[i]\Delta + \mathbf{m}[i]\frac{\Delta}{2}$, d'après Eqn.4.10. Aussi, $\mathbf{y} = Q_\Delta(\mathbf{x} - \mathbf{k}\Delta) - (\mathbf{x} - \mathbf{k}\Delta)$ d'après Eqn.4.11. Évidemment, ces résultats sont donnés pour le cas où l'attaquant reçoit des copies non-distordues comme dans [5].
En analysant Fig.4.3 qui représente la distribution de probabilité du critère de décision $\mathbf{y}[i]$, on remarque que selon la valeur de la clef $\mathbf{k}[i]$ la distribution de probabilité

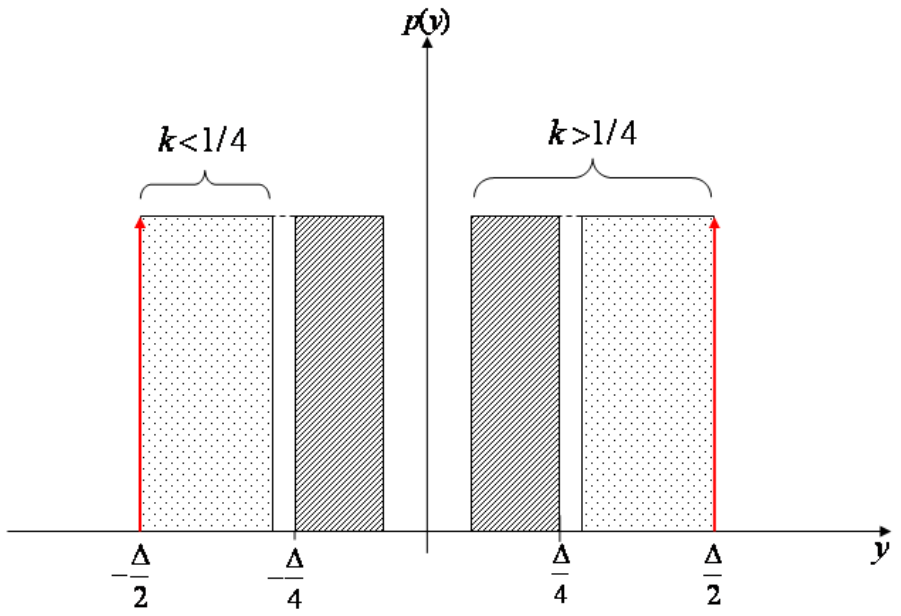


FIGURE 4.3 – La densité de probabilité du critère décision y en fonction des valeurs possibles du signal marqué x dans le cas où le message $m = 1$ dans les cas la valeur de la clef k est (1) inférieure à 1/4 (2) égale à 1/4 (3) supérieure à 1/4.

de la variable critère de décision $\mathbf{y}[i]$ change. Nous obtenons donc 3 cas :

Le cas où aucune clef secrète n'est utilisée ou cas ($\mathbf{k}[i] = 0\ \forall i$) : la densité de probabilité $p(y)$ est la somme de trois Dirac positionnés en $y = 0$, $y = -\frac{\Delta}{2}$ et $y = \frac{\Delta}{2}$. Notons que le premier Dirac correspond au cas où le message inséré est égal 0 et les deux derniers Dirac correspondent au cas où le message inséré est égal à 1.

Le cas où $\mathbf{k}[i] \leq 1/4\ \forall i$: dans ce cas la densité de probabilité représente 2 fenêtres tel qu'il est montré sur Fig.4.3. La largeur de chaque fenêtre ne dépasse jamais $\frac{\Delta}{4}$. Ainsi, l'utilisation d'une clef n'est pas forcément nécessaire si nous utilisons Eqn.4.12.

Le cas où $\mathbf{k}[i] > 1/4\ \forall i$: dans ce cas la largeur des fenêtres dépasse $\frac{\Delta}{4}$, donc, il existe des cas où le décodeur prend des décisions, sur le message inséré, qui sont fausses. Cependant, on note qu'en retranchant à x[i] la valeur $\frac{\Delta}{4}$ la densité du critère de décision reviendrait à celle du cas où $\mathbf{k}[i] \leq 1/4$, autrement dit, la connaissance de la valeur exacte de la clef secrète $\mathbf{k}$ n'est plus nécessaire.

Dans le cas de la QIM avec clef secrète, l'attaquant n'aura pas à estimer la valeur exacte de la clef continue pour accéder au message inséré, mais seulement

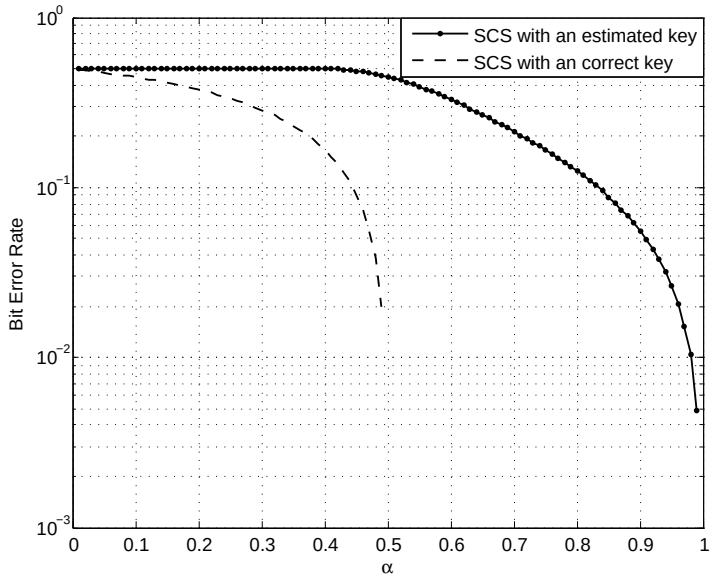


FIGURE 4.4 – Le taux d'erreur binaire en fonction du paramètre α dans le cas d'un système de tatouage SCS avec et sans clef secrète.

déterminer si les valeurs de la clef sont au-dessus ou en-dessous de $1/4$.

Le critère de décision $\mathbf{y\prime}$ que l'attaquant pourrait utiliser pour éviter l'estimation de la clef secrète est formulé comme suit :

$$\mathbf{y'}[i] = \begin{cases} Q_\Delta(\mathbf{x}[i]) - \mathbf{x}[i] \text{ if } |\mathbf{k}[i]| \leq 1/4 \\ Q_\Delta(\mathbf{x}[i] - \Delta/4) - (\mathbf{x}[i] - \Delta/4) \text{ if } |\mathbf{k}[i]| > 1/4 \end{cases} . \tag{4.14}$$

En utilisant Eqn.4.14, nous calculons le taux d'erreurs binaires au niveau du décodeur. Évidemment, nous considérons que l'attaquant sait comment réduire l'estimation de la clef secrète continue en l'estimation des états de cette clef (plus ou moins 1/4) tel que nous l'avons décrit dans la partie précédente. Nous trouvons que le taux d'erreurs binaires du côté attaquant (l'erreur que commet l'attaquant en décodant le message avec Eqn.4.14) est égale à 10^{-6} ! Ceci est évidement très dangereux pour un système de tatouage.

Notons qu'il est possible d'appliquer ce même raisonnement au système de tatouage SCS, avec clef secrète continue, tel qu'il est proposé dans l'article [1]. Cependant, il faudrait que le paramètre de Costa α soit proche de 1 pour que l'attaquant puisse utiliser Eqn.4.14. Fig.4.4 montre que le comportement du SCS sans clef secrète est proche de celui avec clef secrète lorsque Eqn.4.14 est utilisée, puisque le taux d'erreur

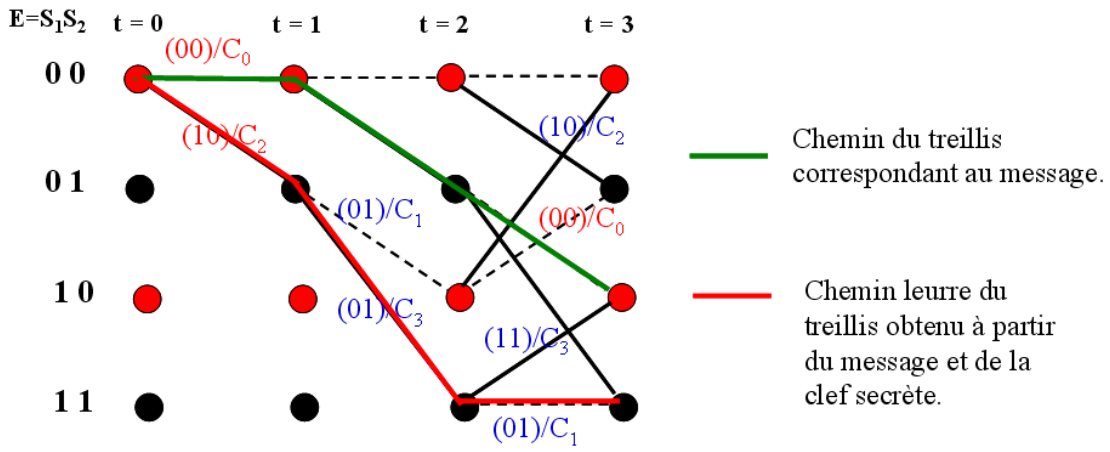


FIGURE 4.5 – Schéma récapitulatif du fonctionnement d'un système de tatouage TCQ sécurisé.

binaire est élevé dans les deux cas pour des valeurs de α faibles (due à un canal de transmission très bruité) mais le b.e.r. est très réduit lorsque α s'approche de 1.
Une simple analyse de la technique qui nous permet de contourner l'estimation de la clef secrète continue montre que le point faible vient du fait que nous utilisons deux dictionnaires uniquement pour insérer le message. Donc, pour sécuriser le système il faudrait trouver une technique qui permet l'insertion en utilisant un nombre de dictionnaires important.

4.3.5 Le tatouage TCQ sécurisé

Dans le but d'éliminer le point faible du système de tatouage QIM, nous proposons d'utiliser la TCQ puisque, d'un côté, le nombre de dictionnaires utilisés est important, de l'autre, l'estimation du message nécessite une connaissance complète du chemin du treillis et donc de la clef ce qui rend la tâche encore plus difficile pour l'attaquant.
Dans la littérature [5] [46], nous trouvons le tatouage TCQ utilisé sans une clef secrète. Généralement, la structure même de la TCQ, basée sur la modulation codée par treillis (treillis coded modulation) [58] [32], permet de protéger le message inséré. Parfois les paramètres du treillis sont maintenus secrets pour protéger le message inséré et donc de jouer le rôle d'une clef secrète.

Afin de sécuriser le tatouage, nous proposons d'utiliser une clef secrète. De ce fait, l'idée ne sera pas de chiffrer uniquement le message mais aussi le chemin du treillis correspondant à ce message. Au même temps, nous nous assurons que les

performances en termes d'invisibilité seront les mêmes qu'un TCQ classique. Ceci est possible grâce au fait que le changement introduit dans le schéma va uniquement modifier les sous-dictionnaires permettant d'établir un chemin dans le treillis correspondant au message à insérer. Ainsi, nous procédons comme suit (voir Fig.4.5) :

1. Nous déterminons d'abord le chemin qui correspond au message que l'on veut insérer, ce qui correspond au chemin représenté par une ligne verte dans Fig.4.5,
2. Nous générons une clef secrète de la même longueur que notre signal hôte ou alors nous utilisons une clef secrète de taille réduite pour générer une séquence pseudo-aléatoire dont la taille est égale à celle du signal hôte,
3. Nous utilisons la séquence pseudo-aléatoire (correspondante à la clef secrète ou pas) pour décaler les valeurs des échantillons quantifiés avec le chemin du message à insérer, i.e. décaller les échantillons hôtes vers un nouveau dictionnaire, donc un nouveau chemin du treillis représenté par la ligne en rouge sur le treillis de Fig.4.5, qui sera transmis à la place du premier.

Pour extraire le message au niveau du décodeur, il suffit d'utiliser la clef secrète pour remettre les échantillons obtenus dans le bon dictionnaire et retrouver ainsi le vrai message inséré au niveau de l'encodeur.

Il est à noter que la clef secrète dans ce cas est une clef discrète dont le nombre d'état est égal au nombre de dictionnaires utilisés dans le treillis. Ceci permet de renforcer la sécurité du système de tatouage. Ce qui rend encore plus la tâche difficile à l'attaquant est le fait qu'il doit tester tout les chemins possibles pour retrouver le message inséré, puisque le chemin correct du treillis doit être estimé dans sa globalité.

Pour évaluer et comparer le niveau de sécurité du schéma de tatouage traité dans ce chapitre, nous utilisons Eqn.4.9 pour déterminer le nombre d'observations (copies marquées) nécessaires pour estimer complètement la clef secrète. Les résultats obtenus sont illustrés sur Fig.4.6.
On constate une amélioration significative du niveau de sécurité du schéma TCQ proposée par rapport au système QIM. Il est clair qu'il sera plus difficile à l'attaquant d'estimer la clef secrète dans le cas de la TCQ sécurisé que dans le cas du QIM sécurisée.
Dans l'article [59] Shannon montre qu'une protection efficace dépend de l'alphabet

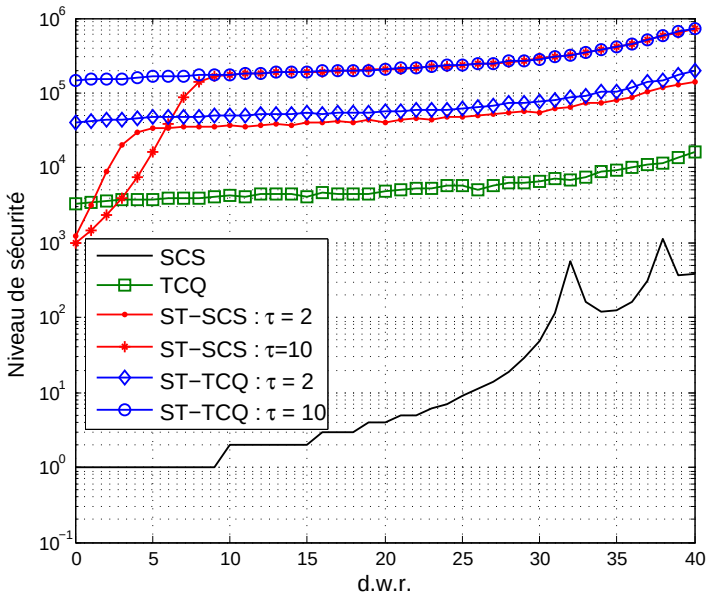


FIGURE 4.6 – Le niveau de sécurité mesuré par le nombre d'observations nécessaire à l'estimation de la clef secrète en fonction du rapport document sur watermark (d.w.r. : document to watermark ratio) pour les systèmes QIM, TCQ, ST-QIM and ST-TCQ.

et de la longueur de la clef secrète. Malheureusement, la taille du système TCQ est limitée au nombre maximum d'états du treillis utilisé. Ainsi, la sécurité du système proposé va certainement être bornée sans pouvoir aller au-delà.

4.3.6 Le ST pour le renforcement du système de sécurité

Il est possible d'améliorer la sécurité du système de tatouage en rajoutant une couche de sécurité supplémentaire. D'un autre côté, l'utilisation du ST permet de choisir une direction secrète pour l'insertion du tatouage numérique sans interférer avec le système de sécurité du tatouage. Ainsi, nous pouvons utiliser le ST comme un système de sécurité alternatif ou alors comme un second niveau de sécurité pour le tatouage numérique.

Nous proposons d'étudier deux combinaisons intéressantes avec le ST. La première est la ST-QIM qui est une version simplifiée du ST-SCS proposé par Eggers et al. dans leur article [1]. D'après Fig.4.6, le ST-QIM a un meilleur niveau de sécurité que la QIM grâce à l'utilisation d'une direction secrète. Cette perfomance est essentiellement obtenue par le fait que le ST permet de choisir une direction secrète dans

un hyper-espace (espace à plusieurs dimensions). Ainsi, le ST peut être vu comme un second niveau de protection pour les systèmes de tatouage.
Une autre combinaison intéressante est le ST-TCQ. Nous l'avions proposée dans l'article [52], où nous avons démontré les bonnes performances de ce système. D'ailleurs, Fig.4.6 montre que le ST-TCQ a le meilleur niveau de sécurité par rapport aux autres systèmes. Ceci est obtenu grâce à la direction secrète fournie par le ST renforcé par la TCQ sécure.

D'un autre côté, la définition de la sécurité est *...l'impossibilité par des utilisateurs non-autorisés d'avoir un accès direct au canal du tatouage* (traduction de la définition donnée par Kalker dans [60] : *Security refers to the inability by unauthorized users to have access to the raw watermarking channel*). Même si l'analyse n'est pas exhaustive, nous avons proposé d'étudier l'accès de l'attaquant au canal du tatouage et son habilité à extraire le message après avoir cassé le système de protection. Les résultats obtenus sont présentés sur Fig.4.7. Ils découlent de la mesure de la quantité de copies N_0 nécessaires à l'estimation complète du message inséré. N_0 est calculé à l'aide de la formule suivante :

$$N_0 = H(M)/I(X;M), \tag{4.15}$$

tel que $H(M)$ représente l'entropie de la variable aléatoire M modélisant le message inséré et $I(X;M)$ est l'information mutuelle entre le signal marqué modélisé par la variable aléatoire X et le message inséré. Notons que cette formule est obtenu exactement de la même manière que Eqn.4.9, sauf que la variable que l'on voudrait estimer n'est plus la clef K mais plutôt le message inséré M.
Généralement, il est supposé que si l'attaquant arrive à estimer la clef alors il aura forcément accès au message pour produire des copies non-autorisées, par exemple. Cependant, Fig.4.7 nous montre que le tatouage selon une direction secrète en utilisant le système ST permet de mettre une autre barrière face l'attaquant pour lire le message. En d'autres termes, la quantité de copies marquées nécessaires pour que l'attaquant puisse lire le message, après l'estimation de la clef secrète et lorsque le ST est utilisé, est beaucoup plus importante que lorsque le ST n'est pas utilisé.

Dans la deuxième partie de ce chapitre, nous nous intéressons à un autre type d'attaques qui sont les attaques par élimination de la marque. De la même façon que pour les attaques par estimation, nous allons analyser et étudier l'effet des attaques

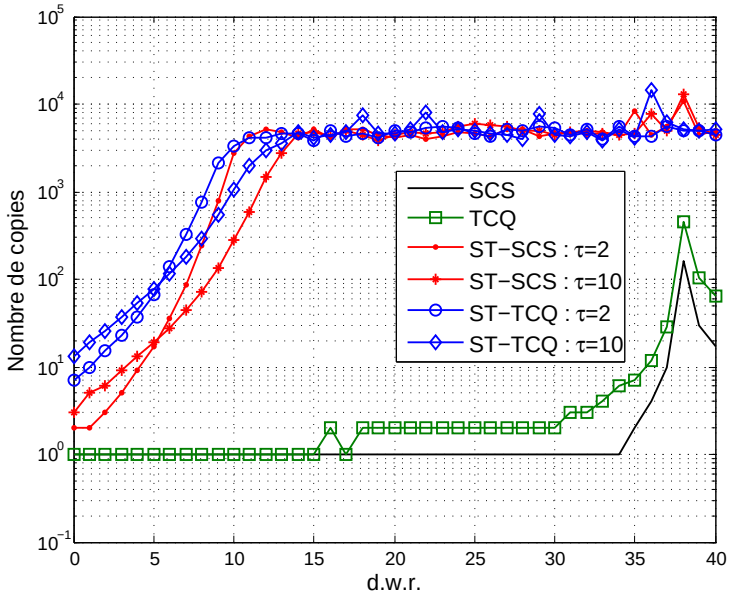


FIGURE 4.7 – La quantité d'observations nécessaire à l'estimation du message watermark inséré en fonction du rapport document sur watermark (d.w.r. : document to watermark ratio) pour les systèmes de tatouage QIM, TCQ, ST-QIM and ST-TCQ.

par élimination de la marque, puis, proposer des solutions pour renforcer la sécurité des systèmes de tatouage. Notons que dans cette thèse, nous nous sommes intéressé, en particulier, à une attaque par élimination qui est l'attaque par moyennage temporelle, où Temporal Frame Averaging (TFA). Ceci est justifié par le fait que c'est une attaque qui permet de faire des simulations sur des signaux réels et les résultats obtenus peuvent être généralisé à d'autres attaques du même type.

4.4 Attaques par élimination de la marque : Attaque TFA

L'attaque par TFA (temporal Frame Averaging) fait partie d'une famille d'attaques appelées : Attaques par élimination de la marque, aussi, elle est parfois classée parmi les attaques par collusion intra document [61]. Elle consiste à enlever le tatouage inséré du document afin d'éliminer toute protection (preuve de propriété ou identifiant de l'utilisateur malicieux).
Tel qu'il a été montré dans l'article [61], les attaques par collusion intra sont des attaques malicieuses et très puissantes [62], puisqu'elles nécessitent l'utilisation d'une

seule copie du signal marqué pour casser le système de sécurité.
Nous présentons dans la suite l'attaque par moyennage temporel (attaque TFA) dans le cas particulier des vidéo. Ce cas a été très étudié [61] [63], puisque l'attaque par moyennage est plus efficace dans le cas de la vidéo que d'autres types de signaux multimédias, à cause de la corrélation temporelle entre les différentes trames et la quantité d'information contenue dans un seul document. Après quelques définitions basiques liées à l'attaque par moyennage, nous présentons l'effet d'un attaque TFA sur une vidéo, puis, nous donnerons les solutions que nous avons développées au cour de la thèse.

4.4.1 Définitions

L'attaque TFA a été bien décrite dans la thèse de Gwénael Doerr [64], elle consiste à calculer la moyenne sur plusieurs trames autours d'une certaine trame vidéo, puis, remplacer celle-ci avec la moyenne. Ceci revient à faire un filtrage passe bas qui permet d'éliminer la marque insérer tout en préservant la qualité visuelle de la vidéo. Ainsi, une attaque TFA sur une fenêtre (nombre de trames) ω qui remplace la trame $\mathbf{x}$ par la moyenne des trames $\dot{\mathbf{x}}$ peut être formulée comme suit

$$\dot{\mathbf{x}} = \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{x}_u, \tag{4.16}$$

tel que $\mathbf{x}_u, u \in [-\frac{\omega}{2}, \frac{\omega}{2}[$, représente l'ensemble des trames voisines de la trame $\mathbf{x}$.

Dans la suite, la vidéo est considérée comme une évolution de L trames à travers le temps. Chaque trame vidéo (ou séquence d'image) est mise sous forme vectorielle que nous notons $\mathbf{s}$ de taille $N \times 1$.

4.4.2 Effet de l'attaque par moyennage sur le Spread Transform (ST)

L'attaque par moyennage est une attaque très puissante contre la plupart des tatouages. Afin de la contrer nous nous somme intéressés à l'utilisation du ST. Ainsi, nous avons considéré dans un premier temps qu'un seul tatouage est inséré pour chaque trame vidéo, comme dans l'article [63]. Donc, le i^{me} échantillon marqué

$\mathbf{x}[i]$ est formulé comme suit,

$$\dot{\mathbf{x}}[i] = \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{s}_u[i] + \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} (\mathbf{x}_u^{\mathrm{st}}[l] - \mathbf{s}_u^{\mathrm{st}}[l]) \cdot \mathbf{t}_u[i]. \tag{4.17}$$

Pour extraire le message inséré, le décodeur utilise le signal reçu donné par la formule suivante :

$$\dot{\mathbf{y}}^{\mathrm{st}}[l] = < \dot{\mathbf{s}}, \mathbf{t} > + \underbrace{\frac{1}{\omega}(\mathbf{x}^{\mathrm{st}}[l] - \mathbf{s}^{\mathrm{st}}[l])}_{Information\ utile} + \underbrace{\frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[, u \neq 0} (\mathbf{x}_u^{\mathrm{st}}[l] - \mathbf{s}_u^{\mathrm{st}}[l]) < \mathbf{t}_u, \mathbf{t} >}_{Interferences} + \underbrace{< \mathbf{v}, \mathbf{t} >}_{Bruit\ projet} \tag{4.18}$$

où,

- $\dot{\mathbf{y}}^{ST}[l]$ représente la l^{th} composante de la transformation du signal attaqué reçu $\dot{\mathbf{y}}$,
- $\dot{\mathbf{s}}$ est la moyenne de ω trames vidéo originales, donnée par,

$$\dot{\mathbf{s}} = \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{s}_u, \tag{4.19}$$

Tel qu'il a été présenté dans les articles [2] [5], en utilisant le fait que la direction d'étalement est normalisée, nous avons,

$$\begin{aligned} \dot{\mathbf{y}}^{ST}[l] &= < \dot{\mathbf{s}}, \mathbf{t} > + \underbrace{\frac{1}{\omega}(\mathbf{x}^{\mathrm{st}}[l] - \mathbf{s}^{\mathrm{st}}[l])}_{Usefull\ Information} + \underbrace{\frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[, u \neq 0} (\mathbf{x}_u^{\mathrm{st}}[l] - \mathbf{s}_u^{\mathrm{st}}[l]) < \mathbf{t}_u, \mathbf{t} >}_{Interferences} \\ &+ \underbrace{< \mathbf{v}, \mathbf{t} >}_{Bruit\ projet} \end{aligned} \tag{4.20}$$

D'après Eqn.4.20, on peut déduire que l'attaque TFA induit des interférences qui diminuent le rapport watermark à bruit (w.n.r. : watermark to noise ratio). Ceci est expliqué par le fait que cette attaque diminue la puissance du tatouage σ_W^2 d'un facteur $1/\omega^2$.

4.4.3 Solution pour contrer l'attaque TFA : exploitation de la diversité temporelle

En se basant sur les résultats de la partie précédente, nous proposons d'utiliser une technique, généralement utilisée en Fingerprinting, pour contrer les attaques par collusions. Elle consiste à utiliser des directions d'insertion du tatouage orthogonal. Nous proposons, en plus, d'exploiter la diversité temporelle de la vidéo. Le but de cette manipulation est de recouvrer les informations perdues en utilisant les trames adjacentes.

Dérivation de l'expression du signal attaqué

Le facteur d'étalement sur les trames vidéo τ_F est considéré de manière à ce que $N\tau_F = \tau$. L'insertion de l'information se fait le long des directions données par les vecteurs $\mathbf{t}$ de taille $(N\tau_F) \times 1$, tel que

$$\mathbf{s}^{\mathrm{st}}[l] = \sum_{i=N\cdot\tau_F\cdot l}^{N\cdot\tau_F\cdot l+N\cdot\tau_F-1} \mathbf{s}[i]\cdot\mathbf{t}\left[i - |[\frac{i}{N}]| + N\cdot l\right] =< \mathbf{s},\mathbf{t} > . \tag{4.21}$$

où $l = |[\frac{i}{N\cdot\tau_F}]|$ tel que $|[.]|$ donne la partie entière de la fraction.
Afin de simplifier la présentation des développements, nous reformulons l'expression donnée par Eqn.4.21 à l'aide du produit scalaire entre deux vecteurs $< ., . >$,

$$\mathbf{s}^{\mathrm{st}}[l] = \sum_{i=N\cdot\tau_F\cdot l}^{N\cdot\tau_F\cdot l+N\cdot\tau_F-1} < \mathbf{s},\mathbf{t} > . \tag{4.22}$$

Aussi, la trame vidéo attaquée par la TFA est donnée par,

$$\dot{\mathbf{y}} = \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[} \mathbf{y}_u, \tag{4.23}$$

où,

$$\mathbf{y} = \mathbf{s} + \mathbf{w}^{\mathrm{st}}[l]\cdot\mathbf{t} + \mathbf{v}, \tag{4.24}$$

représente le signal marqué et transmis à travers le canal de communication.

Dans ce travail, les directions d'étalement ont été choisies mutuellement ortho-

gonales. D'un autre côté, une analyse de l'attaque par TFA indique que son impact est différent selon la localisation des trames vidéo situées au début, au milieu et à la fin de la fenêtre d'attaque.
Comme nous l'avions expliqué au début de ce chapitre sur le ST, l'extraction de l'information dissimulée dans le signal marqué se fait à partir de la transformation du signal reçu $\mathbf{y}^{ST}$. Après des développements et en utilisant quelques propriétés des suites arithmétiques (voir les détails dans l'Annexe B à la fin du manuscrit), nous avons réussi à développer l'expression d'une trame vidéo attaquée, elle est donnée par la formule suivante :

$$\dot{\mathbf{y}}^{\mathrm{st}}[l] = \mathbf{s}^{\mathrm{st}}[l] + \left(\frac{4\tau_F - \frac{7}{4}\omega + 2}{4\tau_F}\right)\left(\mathbf{x}^{\mathrm{st}}[l] - \mathbf{s}^{\mathrm{st}}[l]\right) + \mathbf{v}^{\mathrm{st}}[l]. \tag{4.25}$$

Cette formulation de l'attaque par moyennage sur une vidéo est très importante. Son analyse permet de sortir plusieurs scénarios et plusieurs analyses concernant l'attaque par moyennage sur le ST. Cependant, la principale conclusion est plus l'étalement sur les séquences serait important plus le tatouage inséré sera plus résistant.

Analyse théorique

En procédant à une comparaison entre Eqn.4.20, obtenue dans le cas d'un étalement sur une seule trame, et Eqn.4.25 qui formule notre proposition d'exploitation de la diversité temporelle, on peut déduire ce qui suit :

1. Le signal interférence inter-trames est complètement éliminé grâce à l'utilisation de directions orthogonales,
2. La puissance de tatouage est multipliée par un facteur $(\frac{4\tau_F - \frac{7}{4}\omega + 2}{4\tau_F})^2$ dans Eqn.4.20 d'un facteur $1/\omega^2$ dans un cas classique. Ainsi, dans le cas de l'exploitation de la diversité temporelle, nous avons un degré de liberté de plus puisqu'il nous est possible de fixer le τ_F adéquat pour contrer l'attaque par moyennage.
   L'analyse de la fomule Eqn.4.25 permet de sortir trois situations :
   - ***Situation 1*** **:** dans le cas ou l'attaque par TFA est appliquée avec une fenêtre largement plus grande que l'étalement sur les trames, plus précisément, $\frac{7}{4}\omega \gg 4\tau_F + 2$, donc

$$\dot{\mathbf{y}}^{\mathrm{st}}[l] \approx \left(\frac{\omega}{16\tau_F}\right)\left(\mathbf{x}^{\mathrm{st}}[l] - \mathbf{s}^{\mathrm{st}}[l]\right) + \mathbf{v}^{\mathrm{st}}[l] \tag{4.26}$$

Comme on peut le constater dans cette equation, le signal hôte est complètement perdu. Dans cette situation, le décodeur ne peut extraire l'information mais l'attaquant perd son signal hôte. Donc, l'attaque a échoué.

- ***Situation 2*** **:** Le cas où l'attaque par TFA a une fenêtre d'attaque égale à $\frac{16}{7}\tau_F + \frac{8}{7}$, i.e., $\frac{7}{4}\omega = 4\tau_F + 2$ donc,

$$\dot{\mathbf{y}}^{\mathrm{st}}[l] = \mathbf{s}^{\mathrm{st}}[l] + \mathbf{v}^{\mathrm{st}}[l]. \tag{4.27}$$

Dans ce cas, l'attaque par moyennage est très efficace et enlève complètement le tatouage. Cependant, elle nécessite une connaissance, à priori, du facteur d'étalement. Il suffit de garder ce facteur secret pour sécuriser le système.

- ***Situation 3*** **:** Lorsque l'attaque TFA est appliquée sur un nombre plus petit de trames comparé au facteur d'étalement sur les trames choisies τ_F, plus précisément, $\frac{7}{4}\omega \ll 4\tau_F + 2$, donc,

$$\dot{\mathbf{y}}^{\mathrm{st}}[l] \approx \mathbf{s}^{\mathrm{st}}[l] + (\mathbf{x}^{\mathrm{st}}[l] - \mathbf{s}^{\mathrm{st}}[l]) + \mathbf{v}^{\mathrm{st}}[l] = \mathbf{y}^{\mathrm{st}}[l]. \tag{4.28}$$

Dans ce cas, l'attaque TFA n'a absolument aucun effet sur la marque. Cette situation est la plus probable puisque l'encodeur décide du nombre de trames sur lesquelles l'information est étalée.

La première situation n'a pas lieu d'être en réalité, puisque l'attaquant ne voudrait pas perdre le signal original. Le deuxième scénario nécessite une connaissance, à priori, précise de la valeur du facteur d'étalement sur les trames τ_F. Ceci est très difficile pour l'attaquant puisque ce paramètre est censé rester secret. Donc, seul le troisième scénario est possible en réalité. Si le facteur d'étalement τ_F est choisi assez grand, la fenêtre d'attaque ω ne pourra jamais atteindre $\frac{16}{7}\tau_F + \frac{8}{7}$ pour préserver la qualité visuelle de la vidéo. D'un autre côté, il n'est pas souhaité de prendre des valeurs très grandes du facteur d'étalement sur les trames. Ceci pourrait limiter la capacité du système de tatouage tel qu'il est montré dans l'article [1].

4.4.4 Procédé de génération des directions mutuellement orthogonales

La solution proposée dans cette partie pour contrer l'attaque par moyennage est entièrement basée sur la génération de directions mutuellement orthogonales.
Nous avons proposé d'utiliser la matrice de Walsh-Hadamard [65] qui est une ma-

trice carrée dont les composantes sont proportionnelles à ± 1. Cette matrice est très connue et a été largement étudiée dans la communauté de la communication sans fil [66][67]. D'après la construction de Sylvester [68], la matrice de Walsh-Hadamard, $\mathbf{H}_{2^n}$ d'ordre 2^n où $n \geq 2$ est un entier naturel, peut se construire à l'aide la récurrence suivante

$$\mathbf{H}_{2^n} = \mathbf{H}_2 \otimes \mathbf{H}_{2^{n-1}}, \tag{4.29}$$

tel que $\otimes$ désigne le produit de Kronecker,

$$\mathbf{H}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \tag{4.30}$$

ainsi,

$$\mathbf{H}_{2^{n-1}} = \frac{1}{2^{\frac{n-1}{2}}} \begin{pmatrix} \mathbf{H}_{2^{n-2}} & \mathbf{H}_{2^{n-2}} \\ \mathbf{H}_{2^{n-2}} & -\mathbf{H}_{2^{n-2}} \end{pmatrix}. \tag{4.31}$$

En se basant sur la propriété fondamentale où l'ensemble des fonctions de Walsh forment une base orthogonale dans un intervalle unité, la matrice de Walsh Hadamard est donc une matrice orthogonale, i.e.,

$$\mathbf{H}_{2^n}\mathbf{H}_{2^n}^T = \mathbf{H}_{2^n}^T\mathbf{H}_{2^n} = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}_{(2^n)\times(2^n)}. \tag{4.32}$$

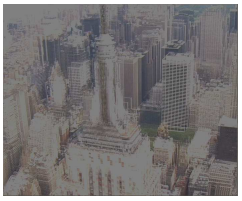
Eqn.4.32 veut dire que les lignes et/ou les colonnes de la matrice $\mathbf{H}_{2^n}\mathbf{H}_{2^n}$ sont mutuellement orthogonales. Donc, chaque direction orthogonale associée à une trames constitue une colonne de la matrice $\mathbf{H}_{2^n}\mathbf{H}_{2^n}$. La vecteur obtenue de la colonne de cette matrice sera répété τ_F fois pour former un vecteur direction orthogonal aux autres directions. Notons que l'ordre de cette dernière est fixé selon les dimensions de la vidéo (signal hôte) $\mathbf{s}$ choisie, i.e., $n = \log_2(N)$. Enfin, le numéro de la colonne choisi peut constituer une clef secrète pour que l'attaquant ne puisse pas retrouver les directions d'insertion.

4.4.5 L'impact visuel de l'attaque TFA

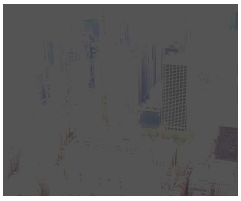
Nous avons mené des expériences d'attaque par moyennage sur un signal vidéo réel, présentées dans Fig.4.8-(a). Pour vérifier les limites perceptuelles de l'attaque

par moyennage nous avons procédé à une attaque de fenêtre $\omega = 2$, présentée sur Fig.4.8-(b), et une fenêtre d'attaque $\omega = 3$ présentée sur Fig.4.8-(c).

(a)



(b)



(c)

FIGURE 4.8 – Impact visuel d'une attaque TFA : (a) Trames vidéo originales (b) Trames dégradées avec une fenêtre d'attaque TFA égale à 2 (c) Trames dégradées avec une fenêtre d'attaque TFA égale à 3

Il est clair que dans le cas du signal vidéo de Fig.4.8-(a) l'attaque TFA introduit des distorsions importantes et dégrade beaucoup la qualité visuelle même pour des fenêtres d'attaques pas très importantes. Ceci n'implique que la marge de manoeuvre en terme de taille de fenêtre d'attaque pour la TFA est très limitée.

4.4.6 Evaluation de la résistance pour une vidéo réelle

Tel qu'il a été indiqué plus haut, nous avons travaillé sur une vidéo presque stationnaire. Ce choix est dicté par le fait que c'est ce genre de vidéo qui favorise les attaques par moyennage du point de vu attaquant. Nous nous mettons, donc, dans un cas difficile côté tatoueur et pourrons évaluer notre solution dans un cas extrême. La vidéo choisie est Bridge (far) qui fait partie des vidéo tests couramment utilisées en traitement des signaux vidéo (voir, par exemple, le site web http ://trace.eas.asu.edu/yuv/index.html). Elle est composée de $L = 2048$ trames de taille 128×128 pixels, ce qui correspond à un $N = 128^2$. Les directions orthogonales

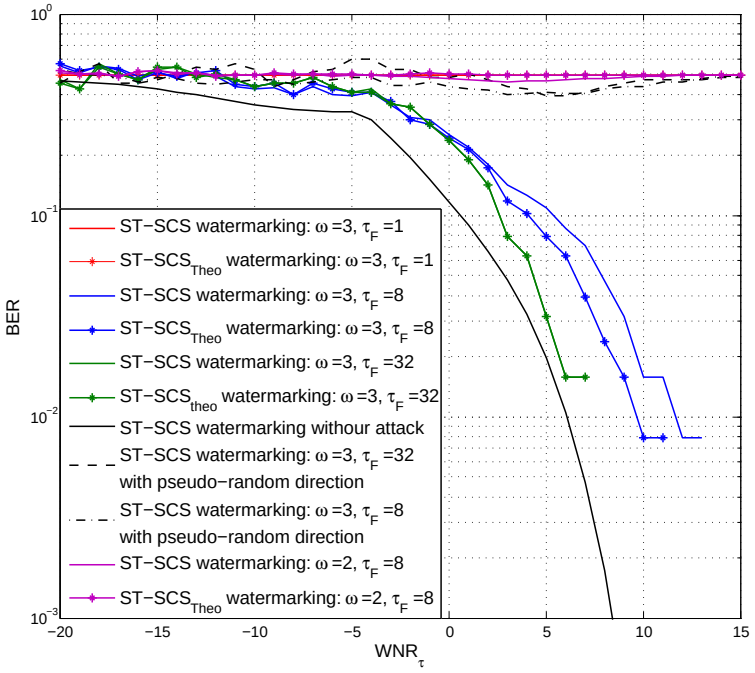


FIGURE 4.9 – Bit error rate (b.e.r.) du ST-SCS avec différents facteurs d'étalement sur les trames τ_F et différentes tailles de la fenêtre d'attaque TFA ω.

t sont obtenues à l'aide de la matrice de Walsh-Hadamard présentée précédemment. La taille de la matrice de Hadamard d'ordre N est de dimension $N \times N$ où chaque colonne nous permet de générer une direction d'étalement composée elle même de τ_F vecteur concaténés. Nous avons constaté que la taille de la matrice de Walsh-Hadamard peut prendre facilement des proportions importantes qui rendent les simulations impossibles. Pour résoudre ce problème nous avons fait appel aux propriétés de la matrice de Hadamard qui permettent une implémentation et une utilisation rapide à faible coût (voir [67] pour plus de détails).

Le but de l'attaquant est d'engendrer le plus de dégâts sur la marque pour obtenir une copie de la vidéo "lessivée" de toute protection. Donc, son but est d'empêcher la lecture de la marque insérée dans la vidéo. Autrement dit, il va tenter d'augmenter le taux d'erreur binaire si l'alphabet du watermark est binaire (ce qui généralement

le cas [1]). Donc, nous avons procédé à l'évaluation du b.e.r. pour mesurer l'efficacité de l'attaque TFA.

Fig.4.9 donne les résultats de nos simulations pour une attaque par moyennage avec différents facteurs d'étalement sur les trames. Notons que lorsque $\tau_F = 1$, ceci correspond à un étalement sur une seule trame vidéo. D'un côté, nous avons ajouté dans Fig.4.9 les courbes obtenues en utilisant les résultats théoriques, en particulier Eqn.4.25, pour les valider. Evidemment, nous avons mis la courbe correspondante au cas sans attaque TFA pour le prendre comme référence et pouvoir visualiser l'effet de cette attaque sur la marque insérée.

D'après Fig.4.9, on constate que le b.e.r. est très élevé dans le cas ou l'étalement se fait sur une seule trame selon des directions orthogonales, ce qui confirme l'analyse théorique précédente de ce cas. Sur la même figure, nous remarquons que lorsque la diversité temporelle augmente, i.e., la valeur de N, donc, τ_F augmente, le b.e.r. décroît sensiblement. Ce qui va dans le sens des nos conclusions théoriques.

Enfin, nous notons que les résultats expérimentaux obtenues sur les images réelles suivent les courbes obtenues des résultats théoriques. Ce qui confirme de plus l'exactitude de nos résultats théoriques, c'est le fait que les courbes du b.e.r. théoriques et expérimentales se superposent parfaitement pour des valeurs du facteur d'étalement τ_F très grandes, tel que le montre la courbe du b.e.r. pour $\tau_F = 32$. Enfin, nous pouvons constater que prendre des valeurs de $\tau_F s$ très grandes permet de réduire considérablement l'effet de l'attaque TFA sur la marque. D'ailleurs, au delà d'une certaine limite nous obtenons la même courbe du b.e.r. que celle du ST-SCS sans attaque. Ceci confirme le scénario 3 donné dans la section précédente.

Dans cette partie, nous avons effectué une analyse théorique de l'attaque par moyennage ou TFA : Temporal Frame Averaging (TFA) lorsque le système ST est utilisé. L'analyse théorique a été proposée dans un cadre général où le ST est utilisé avec n'importe quel système de tatouage. Ensuite nous avons vérifié les résultats obtenus par des simulations sur une vidéo réelle avec le système ST-SCS proposé par Eggers [1].

Dans la suite, nous proposons une solutions généraliste contre les attaques par élimination de la marque. Cette solution exploite les résidus d'un watermark après attaque que nous appelons "cicatrice du tatouage".s

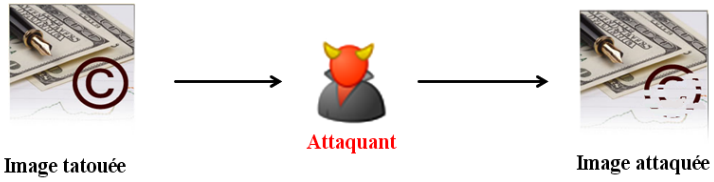
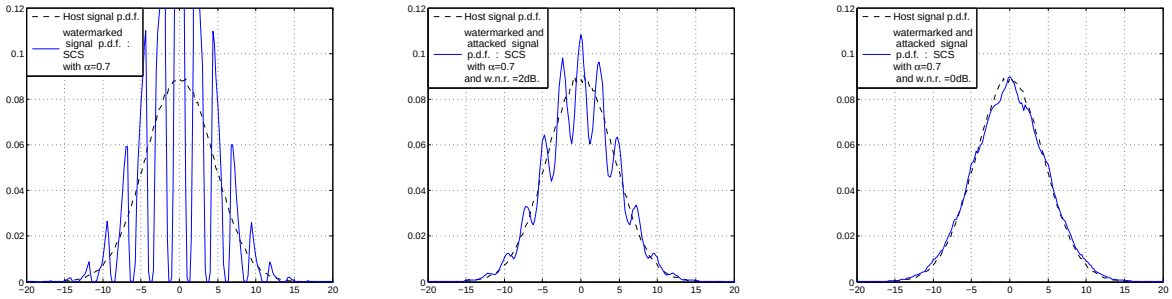


FIGURE 4.10 – Schéma d'attaque par effacement sur un tatouage visible.

4.4.7 Solution pour contrer les attaques par élimination de la marque : utilisation de la cicatrice

Il est connu qu'un système de sécurité ne peut être parfait (à part le masque de Verman en cryptographie connu pour être le système de cryptage incassable [59] qui ne concerne pas notre étude dans cette partie). De même, une attaque aveugle parfaite qui peut "lessiver" le document tatoué ne peut exister à cause des contraintes liées au coût de calcul et l'impact visuel sur le document attaqué, ainsi, il est très probable qu'une attaque par effacement ressemble à l'exemple donné sur la Fig.4.10. Dans cette partie, nous présentons une manière d'exploiter cette imperfection dans les attaques par effacement de la marque. Elle consiste à utiliser les résidus du tatouage après l'attaque que nous appelons ″la cicatrice″ [48] afin de prouver une existence passée d'un tatouage numérique dans le document avant l'attaque par effacement. Ceci permet de cibler les documents sur lesquelles il faudrait pratiquer des post-traitements (généralement coûteux) pour retrouver la marque ou affiner sa détection.
Comme nous l'avons relevé dans le chapitre stéganographie, le problème des systèmes basés sur la quantification est qu'ils distordent les statistiques du signal marqué. Dans cette partie du travail, nous utilisons cette particularité pour mettre en oeuvre le principe de la cicatrice et prouver une existence passée d'un tatouage numérique supposé éliminé par une attaque sur le système. Ainsi, nous évaluons théoriquement les résidus du tatouage numérique, après une attaque par effacement, en utilisant l'information mutuelle entre le signal attaqué et le tatouage sensé être présent avant l'attaque. Ensuite, nous montrons comment il est possible de vérifier, expérimentalement, la quantité de résidus du tatouage encore présente à l'aide de la corrélation normalisée.



(a) (b) (c)

FIGURE 4.11 – Fonction densité de probabilité du signal hôte et marqué utilisant le système SCS pour un rapport document à tatouage (d.w.r. : document to watermark ratio)égal à 13 dB avec différents rapports tatouage à bruit (w.n.r. : watermark to noise ratio) : (a) sans bruit , (b) $w.nr. = 2dB$ and (c) $w.nr. = 0dB$

4.4.8 Définition de la cicatrice du tatouage

Nous définissons la cicatrice du tatouage numérique comme l'information résiduelle d'une marque attaquée. Plus la puissance de la cicatrice est importante, plus grande est la probabilité de détecter le tatouage supposé éliminé. Autrement dit, la cicatrice est le reste d'une information dissimulée après une attaque par effacement. Nous proposons de mesurer cette cicatrice à l'aide de l'information mutuelle $I(X;M)$ entre la copie attaquée modélisée par la variable aléatoire notée X dont les réalisations sont notées x et le message inséré modélisé par la variable aléatoire notée M dont les réalisations sont notées m, tel que,

$$I(X;M) = \sum_{m} \int_{x} p(x,m) log_2(p(x,m)/p(m)), \tag{4.33}$$

où $p(x,m)$ représente la fonction densité de probabilité conjointe entre la copie attaquée et le message inséré.
L'information mutuelle donne une mesure de l'information partagée entre deux variables aléatoires (voir [43] pour plus de détails), ainsi, si l'attaque empêche le décodage à cause d'une élimination partielle du message inséré, on peut espérer qu'il y ait assez de résidus d'information du tatouage pour rendre la preuve d'existence d'un tatouage possible. Ceci est possible si l'information mutuelle entre le message inséré et les copies attaquées dépassent un certain seuil fixé.

4.5 L'interprétation statistique de la cicatrice

Dans le chapitre stéganographie, nous avons développé l'expression de la p.d.f. d'un signal marqué avec le SCS. Nous avons conclu que le tatouage engendre des trous et des bosses dans la distribution de probabilité. L'exemple de Fig.4.11-(a) montre la densité de probabilité d'un signal marqué avec le SCS pour le cas particulier où le paramètre $\alpha = 0.7$.
Dans un contexte de stéganographie passive [5], les discontinuités sur la p.d.f. du signal marqué rendent le système non sécurisé d'après la définition de la sécurité de Cachin [69]. Cependant, les discontinuités sur la p.d.f. du signal tatoué sont une preuve d'existence du tatouage, ce qui est au contraire avantageux dans le cas de détection de la cicatrice. Autrement dit, si le but de l'attaquant est d'éliminer la marque, alors, il est possible d'évaluer l'efficacité de son attaque en analysant les discontinuités sur la p.d.f. du signal tatoué. Ainsi, les Fig.4.11(b) et Fig.4.11(c) montrent l'effet d'une attaque par ajout de bruit AWGN sur les discontinuités de la p.d.f. du signal marqué avec le SCS, ainsi, on peut noter que selon la puissance de l'attaque (puissance du bruit ajouté), les discontinuités sont plus ou moins compensées.
Théoriquement, nous formulons la p.d.f. du signal tatoué et attaqué modélisé par la variable aléatoire Y comme suit :

$$p_Y(y) = \frac{1}{8B\pi\sqrt{\sigma_S^2\sigma_V^2}} \sum_{m,u_m} e^{-\frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}} \cdot \left(erfc(u_m - \frac{\Delta}{2} - C(y)) - erfc(u_m + \frac{\Delta}{2} - C(y)) \right) dz', \quad (4.34)$$

tel que, V est la variable aléatoire modélisant le bruit ajouté de moyenne nulle et de variance σ_v^2, $B = \sqrt{\frac{(1-\alpha)^2\sigma_S^2+\sigma_V^2}{2\sigma_S^2\sigma_V^2}}$, et, $C(y) = \frac{\sigma_S^2(1-\alpha)(y-\alpha u_m)}{(1-\alpha)^2\sigma_S^2+\sigma_V^2}$.

Preuve

$$p_Y(y) = p_X * p_V(y) \quad (4.35)$$

Nous modélisons le signal hôte par l'ensemble des réalisations d'une variable aléatoire Gaussienne, indépendantes et non-stationnaires : S. Ainsi, pour un tatouage SCS, le signal tatoué modélisé par la variable aléatoire X est donné par l'équation suivante :

$$X = (1-\alpha)S + \alpha U, \quad (4.36)$$

où α représente le paramètre d'optimisation de Costa et U représente la quantification de S.
Lorsque l'attaquant procède à un ajout de bruit modélisé par la variable aléatoire V centrée de variance σ_V^2, le signal reçu modélisé par la variable aléatoire Y est donné par la formule suivante,

$$Y = (1-\alpha)S + \alpha U + V. \tag{4.37}$$

Puisque le bruit additif V est indépendant du signal marqué X, la p.d.f. du signal reçu Y est donné par,

$$p_Y(y) = p_X * p_V(y), \tag{4.38}$$

où y représente la réalisation de la variable aléatoire Y.
D'après les articles [1] [70], la p.d.f. du signal marqué est formulée comme suit

$$p_X(x) = \frac{1}{2(1-\alpha)} \sum_{u_m,m} \delta\left(u_m - Q_\Delta\left(\frac{x-\alpha u_m}{1-\alpha}\right)\right) \times p_S\left(\frac{x-\alpha u_m}{1-\alpha}\right), \tag{4.39}$$

donc, la p.d.f. du signal attaqué est donné par la formule suivante,

$$\begin{aligned} p_Y(y) &= \int_{-\infty}^{\infty} \frac{1}{2(1-\alpha)} \sum_{u_m,m} \delta\left(u_m - Q_\Delta\left(\frac{z-\alpha u_m}{1-\alpha}\right)\right) \times p_S\left(\frac{z-\alpha u_m}{1-\alpha}\right) \\ &\cdot \quad p_V(y-z)dz, \end{aligned} \tag{4.40}$$

en procédant au changement de variable suivant : $z' = \frac{z-\alpha u_m}{1-\alpha}$ dans Eqn.(4.40) qui devient,

$$\begin{aligned} p_Y(y) &= \frac{1}{2} \sum_{m,u_m} \int_{-\infty}^{\infty} \frac{1}{2(1-\alpha)} \delta\left(u_m - Q_\Delta\left(z'\right)\right) \times p_S\left(z'\right) \\ &\cdot \quad p_V(y-(1-\alpha)z' - \alpha u_m)dz', \\ &= \frac{1}{2} \sum_{m,u_m} \int_{u_m-\frac{\Delta}{2}}^{u_m+\frac{\Delta}{2}} p_S\left(z'\right) \cdot p_V(y-(1-\alpha)z' - \alpha u_m)dz'. \end{aligned} \tag{4.41}$$

Dans le cas où le signal hôte et le bruit sont considérés Gaussiens, indépendants, identiquement distribués (i.i.d.) avec une moyenne nulle et des variances, égales respectivement, à σ_S^2 et σ_V^2. Donc, la p.d.f. du signal reçu modélisé par la variable

Y donné par Eqn.4.41 peut être formulée comme suit,

$$\begin{aligned} p_Y(y) &= \frac{1}{4\pi\sqrt{\sigma_S^2\sigma_V^2}} \sum_{m,u_m} \int_{u_m-\frac{\Delta}{2}}^{u_m+\frac{\Delta}{2}} e^{-\frac{z'^2}{2\sigma_S^2}} \times e^{-\frac{(y-(1-\alpha)z'-\alpha u_m)^2}{2\sigma_V^2}} dz' \\ &= \frac{1}{4\pi\sqrt{\sigma_S^2\sigma_V^2}} \sum_{m,u_m} \int_{u_m-\frac{\Delta}{2}}^{u_m+\frac{\Delta}{2}} e^{-\frac{z'^2}{2\sigma_S^2}-\frac{(y-(1-\alpha)z'-\alpha u_m)^2}{2\sigma_V^2}} dz'. \end{aligned} \tag{4.42}$$

Nous avons,

$$\frac{z'^2}{2\sigma_S^2} + \frac{(y-(1-\alpha)z'-\alpha u_m)^2}{2\sigma_V^2} = \frac{\sigma_S^2(y-(1-\alpha)z'-\alpha u_m)^2 + \sigma_V^2 z'^2}{2\sigma_S^2\sigma_V^2}. \tag{4.43}$$

Notons que $\sigma_S^2(y-(1-\alpha)z'-\alpha u_m)^2 + \sigma_V^2$ est un polynôme de second ordre, il est possible de l'écrire sous forme de somme de deux carrés :$a(d-cz')+bz'$, tel que : $a=\sigma_S^2$, $b=\sigma_V^2$, $c=(1-\alpha)$ et $d=(y-\alpha u_m)$. Donc,

$$a(d-cz')+bz' = a(d^2-2cdz'+c^2z'^2)+bz' = (ac^2+b)z'^2 - 2acdz' + az'. \tag{4.44}$$

Sachant que,

$$\begin{aligned} ad^2 &= \frac{(ac^2+b)ad^2}{(ac^2+b)} = \frac{a^2c^2d^2}{(ac^2+b)} + \frac{abd^2}{(ac^2+b)} \\ &= \frac{a^2c^2d^2}{(ac^2+b)} + \frac{abd^2}{(ac^2+b)} \end{aligned} \tag{4.45}$$

Eqn.4.44 devient,

$$\begin{aligned} (ac^2+b) & \left[z' - \frac{2abdz'}{(ac^2+b)^2} + \frac{a^2c^2d^2}{(ac^2+b)^2}\right] + \frac{acd^2}{(ac^2+b)} \\ &= (ac^2+b)\left[z' - \frac{abdz'}{(ac^2+b)}\right]^2 + \frac{acd^2}{(ac^2+b)} \\ &= (\sigma_S^2(1-\alpha)^2+\sigma_V^2)\left[z' - \frac{\sigma_S^2(1-\alpha)(y-\alpha u_m)}{(\sigma_S^2(1-\alpha)^2+\sigma_V^2)}\right]^2 \\ &+ \frac{\sigma_S^2\sigma_V^2(y-\alpha u_m)}{(\sigma_S^2(1-\alpha)^2+\sigma_V^2)}. \end{aligned} \tag{4.46}$$

Donc,

$$\frac{z'^2}{2\sigma_S^2} + \frac{\left(y-(1-\alpha)z'-\alpha u_m\right)^2}{2\sigma_V^2} = \left(B(\tau' - C(y))\right)^2 + \frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}, \quad (4.47)$$

tel que,

$$B = \sqrt{\frac{(1-\alpha)^2\sigma_S^2+\sigma_V^2}{2\sigma_S^2\sigma_V^2}},$$

et,

$$C(y) = \frac{\sigma_S^2(1-\alpha)(y-\alpha u_m)}{(1-\alpha)^2\sigma_S^2+\sigma_V^2}.$$

Donc, Eqn.4.47 devient,

$$\begin{aligned} p_Y(y) &= \frac{1}{4\pi\sqrt{\sigma_S^2\sigma_V^2}} \sum_{m,u_m} \int_{u_m-\frac{\Delta}{2}}^{u_m+\frac{\Delta}{2}} e^{\left(-(B(z'-C(y)))^2 - \frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}\right)} dz' \\ &= \frac{1}{4\pi\sqrt{\sigma_S^2\sigma_V^2}} \sum_{m,u_m} e^{-\frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}} \int_{u_m-\frac{\Delta}{2}}^{u_m+\frac{\Delta}{2}} e^{-(B(z'-C(y)))^2} dz' \quad (4.48) \end{aligned}$$

En procédant au changement de variable suivant : $t = (B(z'-C(y)))$, Eqn.4.48 devient,

$$\begin{aligned} p_Y(y) &= \frac{1}{8B\pi\sqrt{\sigma_S^2\sigma_V^2}} \\ &\cdot \sum_{m,u_m} e^{-\frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}} \frac{2}{\sqrt{\pi}} \int_{u_m-\frac{\Delta}{2}-C(y)}^{u_m+\frac{\Delta}{2}-C(y)} e^{-t^2} dz' \\ &= \frac{1}{8B\pi\sqrt{\sigma_S^2\sigma_V^2}} \\ &\cdot \sum_{m,u_m} e^{-\frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}} \left(\frac{2}{\sqrt{\pi}} \int_{u_m-\frac{\Delta}{2}-C(y)}^{\infty} e^{-t^2} - \frac{2}{\sqrt{\pi}} \int_{u_m+\frac{\Delta}{2}-C(y)}^{\infty} e^{-t^2} \right) dz', \quad (4.49) \end{aligned}$$

donc,

$$p_Y(y) = \frac{1}{8B\pi\sqrt{\sigma_S^2\sigma_V^2}} \cdot \sum_{m,u_m} e^{-\frac{(y-\alpha u_m)^2}{2((1-\alpha)^2\sigma_S^2+\sigma_V^2)}} \left(erfc(u_m - \frac{\Delta}{2} - C(y)) - erfc(u_m + \frac{\Delta}{2} - C(y)) \right) dz', \tag{4.50}$$

Heureusement, l'attaquant ne peut augmenter la puissance du bruit additif indéfiniment puisqu'il doit tenir compte de l'impact visuel qu'aura son attaque sur la vidéo. Tel qu'il est montré sur Fig.4.12, le tatouage est imperceptible sur l'image marqué comme montré sur la Fig.4.12(a). Cependant, la Fig.4.12(b) montre que la marque est bien visible sur la p.d.f. de l'image marquée. Aussi, lorsque l'attaquant procède à une attaque AWGN tel que le w.n.r. est égale à 5 dB, on note sur la Fig.4.12(c) que la qualité de l'image reste acceptable mais des distorsions dues au tatouage sur la p.d.f. restent bien présentes comme le montre Fig.4.12(d). C'est exactement ce reste des effets du tatouage que l'on nomme cicatrice et qui peut être utilisés comme preuve d'une existence passée d'une marque, même si celle-ci ne peut plus être décodée par le récepteur. Dans le cas où l'attaquant procède à une attaque beaucoup plus puissante, tel que le $w.n.r.$ est égal à -5 dB, nous constatons sur la Fig4.12(f) que la p.d.f. de l'image devient lisse (sans distorsions dues au tatouage numérique). Cependant, la qualité de l'image est fortement dégradée comme le montre Fig.4.12(e). Ceci montre que l'attaquant est limité par la distorsion visuelle induite par son attaque.

4.6 La cicatrice du tatouage numérique en pratique

Comme nous l'avons vu dans le point précédent, la cicatrice du tatouage peut être un moyen puissant et plus le système de tatouage laisse des traces, même transparentes pour l'utilisateur, lors de l'insertion de la marque, plus il y a de chance de retrouver la marque et d'améliorer la protection des copies diffusées.
Comme précisé précédemment, la cicatrice du tatouage numérique peut être mesurée à l'aide de l'information mutuelle entre le tatouage inséré et le signal marqué

attaqué. Ainsi, nous avons calculé cette information mutuelle pour deux systèmes de marquage : le SCS et la QIM, comme montré sur Fig.4.13(a). Comme prévu, la puissance de la cicatrice diminue lorsque la puissance de l'attaque augmente. Cependant, on note que la puissance de la cicatrice dans le cas du SCS est un peu plus importante que celle de la QIM. Ceci est dû au fait que le tatouage SCS est plus robuste que la QIM, ce qui préserve mieux le signal marqué contre l'attaque AWGN. D'un autre côté, sur la Fig.4.13(b), nous avons calculé la corrélation entre le signal tatouage attaqué donné par $\dot{\mathbf{w}}$ et le signal tatouage inséré $\mathbf{w}$. Dans ce travail de thèse, nous avons choisit de calculer la similarité entre les deux signaux à l'aide de la corrélation normalisée donnée par la formule suivante,

$$Correlation = \frac{\sum_i (\dot{\mathbf{w}}[i] - \bar{\dot{w}})(\mathbf{w}[i] - \bar{w})}{\sqrt{\sum_i (\dot{\mathbf{w}}[i] - \bar{\dot{w}})^2} \sqrt{\sum_i (\mathbf{w}[i] - \bar{w})^2}}, \tag{4.51}$$

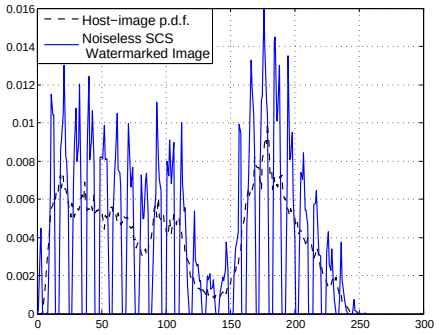
où $\dot{\mathbf{w}}[i]$, $\bar{\dot{w}}$, $\mathbf{w}[i]$ et $\bar{w}$ sont, respectivement, la i^{eme} composante du signal tatoué attaqué, la moyenne du signal attaqué, la i^{eme} composante du signal tatoué non-attaqué et la moyenne du signal tatoué non-attaqué.
D'après Fig.4.13(b), la similarité entre le signal tatouage, dans le cas de la QIM, et le signal tatouage attaqué est une fonction décroissante de la puissance d'attaque, comme la cicatrice de la marque. Ainsi, en fixant un seuil au-delà duquel une présence de la marque est confirmée, il nous est possible de récupérer notre tatouage attaqué en procédant à un post-traitement par exemple.

4.7 Conclusions

Dans ce chapitre, nous avons traité du tatouage numérique, tout particulièrement, de la sécurité des systèmes de tatouage. L'étude de la sécurité a conserné deux classes d'attaques : les attaques par effacement et les attaques par estimation. Pour les attaques par estimation, nous avons proposé une analyse théorique de la sécurité de certains systèmes de tatouage informés. Ensuite, nous avons proposé une solution qui consiste à modifier un système existant appelé la TCQ et en le combinant avec le ST considéré dans ce cas comme une couche qui sert à renforcer la sécurité. Pour les attaques par effacement, nous avons proposé une étude théorique de la resistance du ST face à l'attaque TFA sur une vidéo. Ainsi, nous avons pu proposer une implémentation pratique du ST pour qu'il puisse résister au moyennage temporel sur les frames vidéos. Ensuite, la notions de cicatrice a été introduite pour

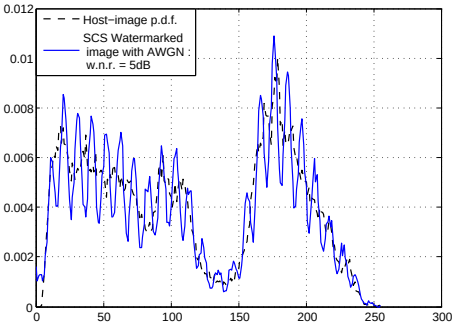
contrer les attaques par effacement de manière générale.
Dans le chapitre suivant, une implémentation pratique d'un schéma basé sur la quantification sera proposée pour une utilisation sur un flux compressé du standard H.264.

(a)



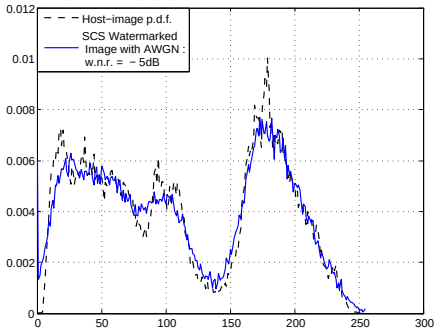
(b)

(c)



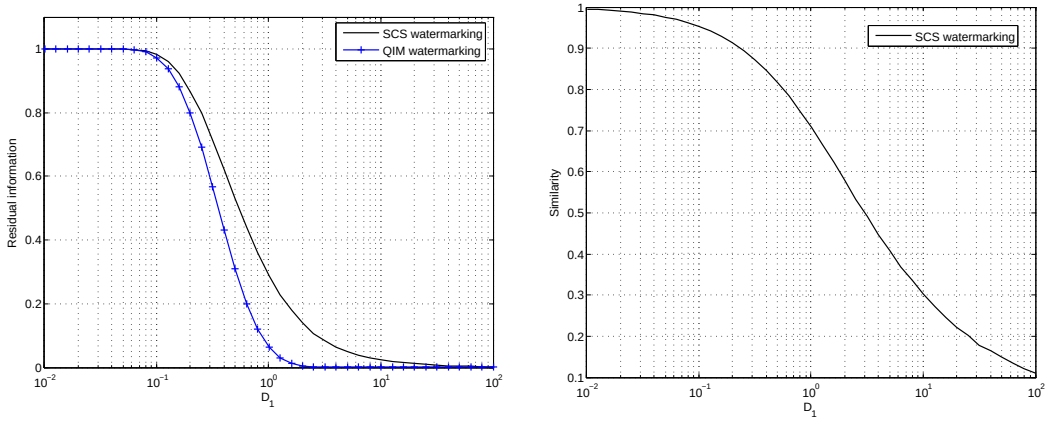
(d)

(e)



(f)

FIGURE 4.12 – L'interprétation de la cicatrice dans le cas d'images réelles lorsque le rapport document à watermark (d.w.r.) est égale à $13dB$: (a) image tatouée avec le système QIM, (b) l'histogramme de l'image originale et de l'image tatouée avec la QIM, (c) image tatouée avec le système QIM et attaquée tel que lorsque le rapport watermark à bruit (w.n.r.) est égale à $5dB$, (d) l'histograme de l'image originale et de l'image attaquée/tatouée avec la QIM lorsque le w.n.r. est égale à $5dB$, (e) image tatouée avec le système QIM et attaquée tel que lorsque le w.n.r. est égale à $-5dB$ et (f)l'histograme de l'image originale et de l'image attaquée/tatouée avec la QIM lorsque le w.n.r. est égale à $-5dB$.



(a) (b)

FIGURE 4.13 – (a) L'information résiduelle (cicatrice) du tatouage inséré après une attaque AWGN en fonction de la puissance du bruit ajouté D_1 (b) la similarité donnée par la corrélation normalisée entre le signal tatouage attaqué et le signal tatouage en fonction de la puissance du bruit ajouté D_1.

Chapitre 5

Conclusions et Perspectives

Cette thèse traite de l'étude des systèmes informés de dissimulation de données. L'objectif est de proposer des améliorations sur les systèmes existants ainsi que des formulations théoriques des comportements de ses systèmes dans différents contextes. Ce manuscrit de thèse s'articule autour de trois grands thèmes :

- La stéganographie passive et active.
- Le tatouage numérique robuste.
- La dissimulation d'information dans un flux compressé vidéo H.264.

Dans un premier temps, nous avons démontré que les systèmes informés basés sur la quantification pouvaient être utilisés en stéganographie. Ainsi, nous avons prouvé que les schémas SCS (Scalar Costa Scheme) et la TCQ (Trellis Codec Quantization) permettent de transmettre un stégo-message avec un niveau acceptable de stégo-sécurité.

De même, le ST (Spread Transform) procure à ces systèmes un bon niveau d'indetectabilité grâce à ces propriétés de projection et d'étalement sur le signal hôte. Plusieurs théorèmes ont été proposés et prouvés concernant les densités de probabilité dans un contexte de stéganographie. L'étude a été complétée par l'analyse du compromis robustesse-capacité-invisibilité pour différents systèmes informés, où il apparait que la technique ST-TCQ, proposée dans cette thèse, est celle qui offre le meilleur compromis.

Dans un deuxième temps, nous nous sommes intéressés au contexte du tatouage numérique et particulièrement à l'aspect sécurité. Ainsi, deux types d'attaques ont été traitées : les attaques par estimation et les attaques par élimination de la marque. Pour le premier type d'attaques, nous avons évalué théoriquement la sécurité de la

QIM (Quantization Index Modulation) et celle du SCS. Ces évaluations ont été validées par des simulations numériques.

Nous avons proposé par la suite, une nouvelle implémentation du TCQ qui permet de mieux résister aux tentatives d'estimation de la clef et du message. Cette nouvelle version de la TCQ combinée avec le ST permet d'améliorer encore plus le niveau de sécurité du tatouage numérique.

De même, nous avons proposé un nouvelle implémentation du ST à partir des formulations théoriques lui permettant de mieux résister à l'attaque TFA (Temporal Frame Averaging).

Ensuite, la notion de cicatrice du tatouage attaqué a été introduite afin de récupérer ou de détecter la présence des résidus d'un tatouage numérique partiellement éliminer. Cette solution permet ainsi d'agir en aval pour mettre en échec les attaques contre le tatouage.

Les éléments étudiées au cours de la thèse ouvrent de nouvelles perspectives pour l'analyse et l'étude des systèmes de dissimulation d'information. Ainsi, nous proposons quelques perspectives aux travaux présentés dans ce manuscrit :

- Appliquer les développements proposés pour la stéganographie à d'autres systèmes, tels que ceux utilisant les lattices.
- Élargir l'étude du comportement des systèmes informés basés sur la quantification à la stéganographie avec gardien malicieux.
- Appliquer les résultats théoriques pour la sécurité face aux attaques par estimation en utilisant l'analyse en composantes principales.
- Appliquer la solution proposée pour contrer l'attaque TFA dans un cadre de fingerprinting destiner à la protection des multimédias dans les réseaux peer to peer.
- Utiliser la notion de cicatrice avec des outils de post processing plus puissant que la similarité, tels que le débruitage ou l'analyse en composantes indépendantes.
- Développer un schéma d'insertion équivalent à celui proposé dans ce manuscrit pour un flux compressé H.264 dans le cas d'un codage entropique CABAC.

Enfin et dans le but de donner une dimension pratique aux travaux de cette thèse, nous avons proposé dans l'annexe B de ce manuscrit d'appliquer certains systèmes étudiés sur la vidéo compressée avec le standard H.264, dans le cadre du projet MEDIEVALS de l'agence nationale de la recherche (ANR). Ainsi, nous avons

proposé un schéma d'implementation des techniques de tatouage informés sur un flux vidéo compressé. Après une étude du standard vidéo MPEG-4/AVC/H.264, un schéma d'insertion dans les coefficients DCT d'un flux compressé H.264 a pu être développé. Les résultats préliminaires présentés dans ce manuscrit montrent que le schéma d'insertion est réalisable et peut être utilisé pour des applications industriel pour un large public d'utilisateurs.

Chapitre 6

Appendix A
Application de la dissimulation des données à un flux vidéo compressé au standard H.264

Cet annexe présente une étude pratique sur la protection des vidéos en utilisant les techniques de data hiding. Sachant que les vidéos sont souvent échangées sous format compressés, nous nous somme intéréssé au standard de compression H.264. Ce dernier a déja été utilisé pour sécuriser l'échange des fichiers vidéo [71] [72] [73] puisque 66% des vidéos échangé sur internet sont sous format H.264 [74].
Dans cet annexe, une présentation du contexte dans lequel ce travail a été effectué est donnée. Ensuite, le standard H.264 est décrit ainsi que ses composantes. Enfin, une présentation d'une méthode d'insertion d'information dans un flux H.264 avec quelques résultats obtenus durant la thèse sont donnés.

6.1 Contexte du travail

Les travaux présentés dans cet annexe s'inscrivent dans le cadre du projet MEDIEVALS financé par l'Agence Nationale de Recherche (ANR). Ce projet a été lancé en 2008 dont l'ambition etait de fournir à l'industrie audiovisuelle une solution, dans le respect des contraintes de normalisation (standard de compression MPEG4/H.264 pour la vidéo et AAC pour l'audio, par exemple) de protection de flux multimédias

de bout à bout, tout en prenant en compte la contrainte transactionnelle (personnalisée en fonction de l'utilisateur final).
Il a été donc proposé d'associer une technique de chiffrement sélectif du flux compressé avec une solutions de tatouage/fingerinting pour des contenus multimédia (audio et vidéo). Le schéma global du système proposé est donné dans Fig.6.1.
Les avantages d'une telle solution par rapport aux systèmes classiques d'embrouillage sont :

- à aucun moment, les utilisateurs non-authorisés ne peuvent avoir accès aux contenus désembrouillés, puisque les opérations de déchiffrement et de l'extraction de la marque ne sont pas corrélées et se font en deux étapes bien distinctes,
- il est possible de sécuriser des contenus audio ou vidéo très volumineux sans augumenter significativement la complexité de l'opération. Ceci est possible grâce aux techniques choisies, qui permettent un traitement rapide des données et surtout grâce au fait qu'une grande partie des opérations (essentiellement celles qui nécessitent des calculs lourds comme la recherche des localisations de la marque) s'effectuent au niveau du serveur et le reste s'effectue au niveau du client. Ainsi, les parties qui sont traitées au niveau serveur permettent, entre autre, d'insérer les informations globales, liées au propriétaire du contenu par exemple. Dans ce cas, il est possible de distribuer les parties volumineuses du contenu via des réseaux de super-distribution plus connus sous le nom de distribution Pair-à-pair (Peer-to-peer).

Technologie medialive ou medialiving

Pour développer le système de protection des données multimédia du projet MEDIEVALS, il a été décidé de fusionner le système de sécurisation des échanges vidéo, dévellopés par l'entreprise medialive, avec les schémas de tatouage numérique. Le système de sécurisation des échanges multimédia ou le medialiving peut être résumé dans les points suivants :

1. La vidéo est considérée sous sa forme compressée,
2. Le flux binaire est analysé pour repérer les parties les plus pertinantes et qui peuvent être manipulées sans incidence sur le processus de décompression,
3. Les parties importantes du flux MPEG-4/H.264 appelées *Contrôle Object (CO)* sont extraites puis remplacées par d'autres éléments appelés *Leurres* et qui

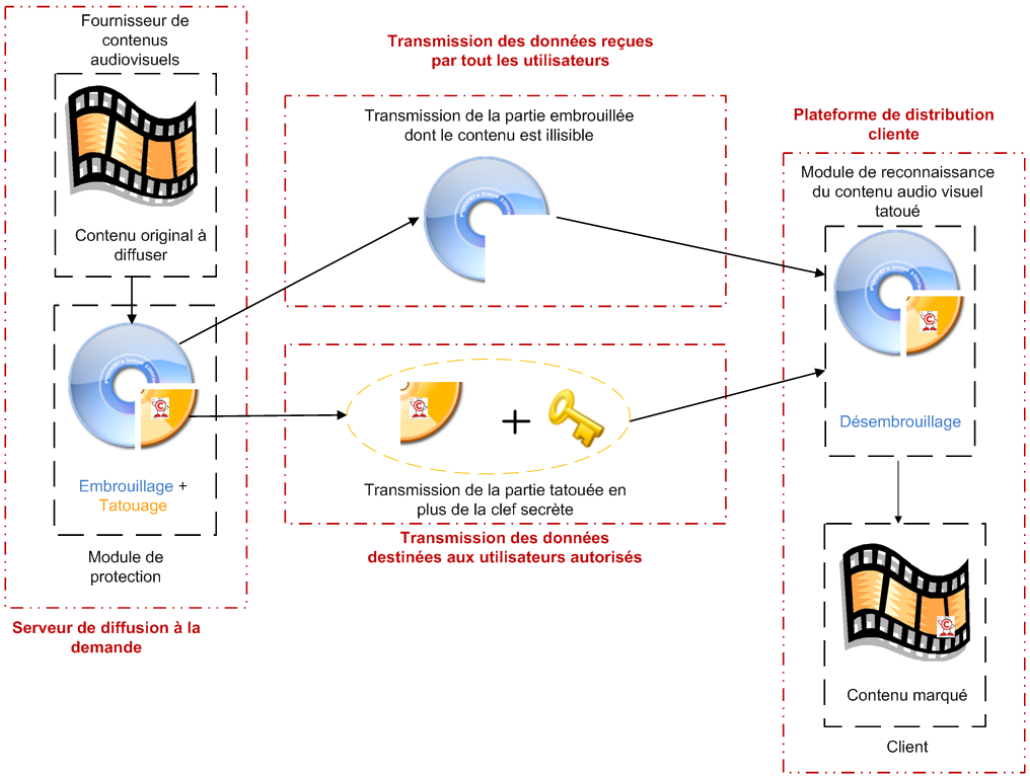


FIGURE 6.1 – La quantité d'observations nécessaire à l'estimation du message watermark inséré en fonction du rapport document sur watermark (d.w.r. : document to watermark ratio) pour les systèmes de tatouage QIM, TCQ, ST-QIM and ST-TCQ.

rendent la lectures de la vidéo impossible, i.e. la dégrade fortement,

4. Le flux leurré et le CO crypté sont transmis séparement.

Ainsi, seuls les utilsateurs autorisés reçoivent la clef de décryptage pour extraire le CO et lire correctement la vidéo. Autrement dit, le CO est décrypté au niveau de l'utilisateur et ses élements remis à leur place d'origine pour restaurer la vidéo. Notons que les travaux présentés dans ce chapitre concernent essentiellement la partie insertion de l'information. De plus, les résultats obtenus représentent qu'une partie de ceux attendus, étant donné que le projet MEDIEVALS est toujours en cours au moment de la rédaction de ce manuscrit.

6.2 Compression H.264 et dissimulation de l'information

Le tout premier standard proposé par l'ITU-T (*International Telecommunications Union - Telecommuniction*), H.120 [75], met en oeuvre cette idée. Dans cette norme, les images sont découpées en blocs. Les blocs identiques et leurs correspondants dans l'image précédente ne sont pas codés et les autres sont traités en mode Intra.

Les codeurs de type H.120 sont cependant inefficaces en cas de mouvement d'ensemble de la caméra vis-à-vis de la scène filmée. Pour résoudre ce problème, la solution a été d'exploiter une partie des informations contenues dans l'image précédente pour prédire les blocs de l'image courante et de transmettre des données permettant de corriger cette prédiction [76]. Ce principe a donné naissance aux codeurs vidéo hybrides, dont le nom provient du fait qu'ils utilisent deux techniques de réduction de redondances : d'une part, une prédiction temporelle, d'autre part, une transformation des résidus de prédiction. Cette structure de base a été formalisée par l'ITU-T dans le standard H.261 [77]. Les standards ultérieurs, MPEG-1 [78], MPEG-2 [79], H.263 [80], MPEG-4 Part 2 Visual [81] et H.264/AVC [82] ont essentiellement repris et amélioré (fortement) cette structure de base de codage hybride.

Pour sécruriser la transmission vidéo, nous avons adopté la norme H.264/MPEG-4 comme standard de compression vidéo. Ceci est justifié par le fait que cette norme est en train de remplacer toute les normes existantes, surtout pour les échanges des contenus multimédia qui demande un niveau de robustesse élevé. D'un autre côté,

le medialiving et la pluspart des produits étudiés et proposés par les partenaires du projet sont basés sur cette norme de compression vidéo.

6.2.1 Couche codage vidéo (VCL : Video Coding Layer)

L'image d'entrée est divisée en macroblocs. Chaque macrobloc est composé de trois composantes : Y, U et V. Du fait que la vision humaine soit moins sensible à la chrominance qu'à la luminance, les macroblocs de chrominance sont sous-échantillonnés d'un facteur 2 dans les directions horizontales et verticales. Par conséquent, chaque portion élémentaire de l'image est composée d'un macrobloc de luminance de 16x16 pixels et de deux macroblocs de chrominance de 8x8 pixels. Chaque macrobloc est prédit en mode Intra ou Inter. Ces deux outils caractérisent l'étage de prédiction du codeur. Le mode Intra exploite les redondances spatiales des images, il permet de construire une estimation d'un macrobloc en utilisant exclusivement les informations contenues dans l'image courante. Le mode Inter tire parti des redondances temporelles entre les images, il permet de prédire le macrobloc courant en utilisant les informations contenues dans des images de référence, qui ont déjà été encodées, décodées puis stockées dans une mémoire. Ce principe de compensation de mouvement repose sur l'estimation d'un vecteur de déplacement associé à chaque bloc. Ce vecteur caractérise la position du bloc le plus vraisemblable dans l'image de référence. Il est évident que la prédiction Inter est beaucoup plus efficace que la prédiction Intra car les redondances temporelles représentent une forte proportion de l'énergie du signal. Le mode Inter est donc utilisé en priorité dans les codeurs vidéo, excepté dans la situation où la mémoire ne contient aucune image de référence (première image d'un film par exemple). L'erreur de prédiction, qui correspond à la différence entre le bloc original et le bloc prédit, est ensuite transformée, quantifiée puis compressée par le biais d'un codage entropique.

Les informations de contrôle ainsi que les vecteurs de mouvement sont également compressés avant d'être transmise. De manière à reconstruire les mêmes images à l'encodeur et au décodeur, les coefficients quantifiés de chaque bloc sont inversés et ajoutés au signal de prédiction. Cette opération permet de reconstruire chaque macrobloc compressé, qui pourra ensuite servir de référence pour la prédiction des autres macroblocs. Afin de réduire les artefacts entre les blocs, un filtre débloquant a été intégré dans la boucle de compression. Ce dernier permet de lisser les informations visuelles avant de les stocker dans la mémoire de référence. La structure de

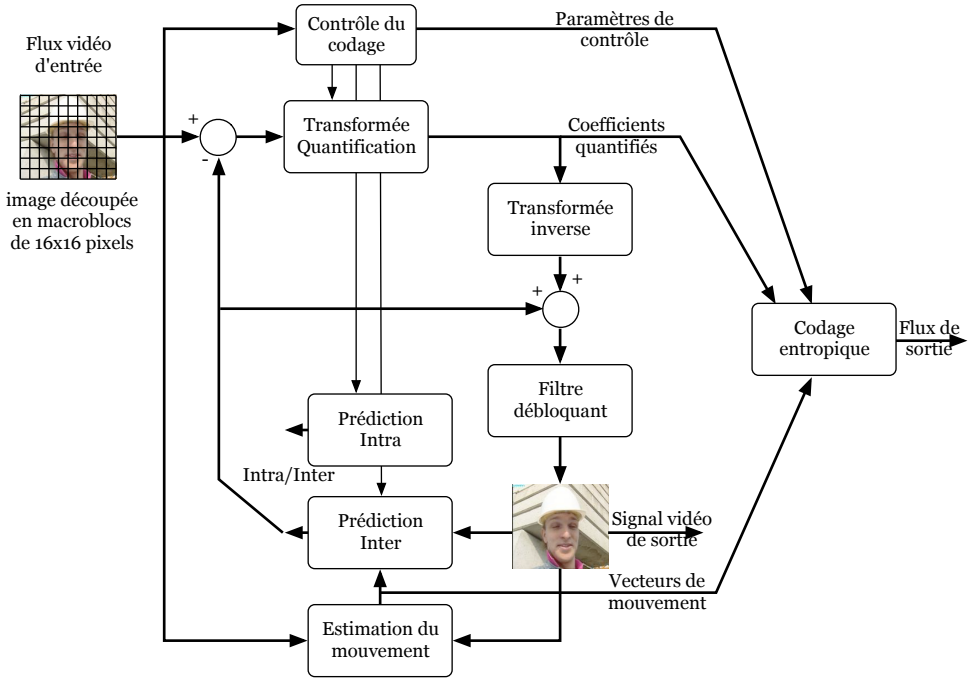


FIGURE 6.2 – Principe du codage H.264.

traitement du décodeur est plus simple que celle de l'encodeur. Le décodeur commence par décompresser les différents types d'informations : paramètres de contrôle, vecteurs de mouvement et coefficients quantifiés. Les macroblocs sont ensuite successivement prédits en utilisant le mode approprié (Intra ou Inter). En parallèle, les résidus de prédiction sont reconstruits grâce aux coefficients quantifiés, puis ajoutés au signal de prédiction. Suite à l'opération de filtrage, le macrobloc est complètement décodé et peut être stocké en mémoire pour les prédictions futures. Un schéma récapitulatif du standard H.264 est présenté sur la Fig.6.2.

Prédiction inter-image

La prédiction inter-image créé un modèle prédictif (les macroblocs P) à partir d'une ou plusieurs frames vidéo précédemment codées (macroblocs I). Notons que le H.264 utilise un large support de taille de blocs (de 16x16 à 4x4) et un sous-échantillonnage plus fin des vecteurs de mouvement (résolution au quart d'échantillon pour la luminance), d'où une complexité calculatoire accrue. Cette prédiction est réalisée par compensation de mouvement des blocs. Des vecteurs de mouvement sont prédits et calculés suivant un certain découpage de l'image.

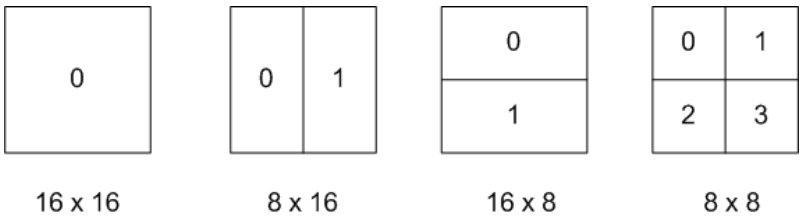


FIGURE 6.3 – Partitions de macroblocs : 16x16, 8x16, 16x8, 8x8.

Compensation de mouvement Chaque macrobloc 16x16 peut être découpé de quatre manières différentes (voir Fig.6.3) et la compensation de mouvement peut être appliquée à une partition 16x16, à deux partitions 8x16, ou à quatre partitions 8x8. Si le mode 8x8 est choisi, les quatre sous-macroblocs peuvent à leur tour être découpés (Fig.6.4) en une partition 8x8, deux partitions 8x4, deux partitions 4x8 ou quatre partitions 4x4. Ces partitions et sous-macrobloc donnent accès à un grand nombre de combinaisons pour chaque macrobloc. Cette méthode est connue sous le nom de compensation de mouvement à structure d'arbre.

Chaque partition ou sous-macrobloc nécessite un vecteur de mouvement. Chaque vecteur doit être codé et transmis et le choix du découpage retenu doit également être inclus dans le bitstream. Choisir une grande taille de partition (16x16, 16x8, 8x16) entraîne un nombre de bits réduit pour signaler le choix des vecteurs de mouvement et le type de partition.

Par contre, le résidu obtenu peut contenir une grande quantité d'énergie dans les zones d'image très détaillées. Une petite taille de partition (8x4, 4x4, . . .) entraîne un résidu de faible énergie mais nécessite un grand nombre de bits pour coder les vecteurs et le choix de partition. De plus, le choix de la taille de partition a un impact significatif sur les performances de compression. De manière générale, une grande partition est plus appropriée pour des zones homogènes et une petite pour des zones de détails. Chaque composante de chrominance d'un macrobloc (Cb et Cr) a une résolution moitié moindre que la luminance. Chaque bloc de chrominance est partitionné de la même manière que la luminance, avec une taille de partition divisée par deux horizontalement et verticalement (une partition 8x16 de luminance correspond à une partition 4x8 de chrominance). Les composantes horizontales et verticales de chaque vecteur de mouvement sont ensuite dédoublées pour les blocs de chrominance.

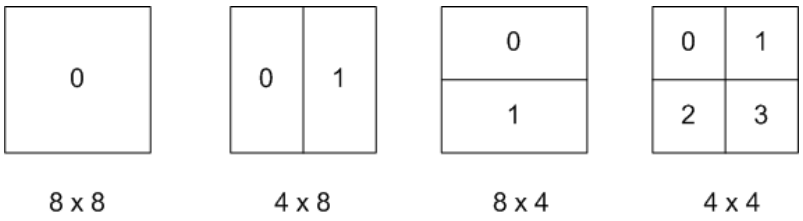


FIGURE 6.4 – Partitions de macroblocs : 8x8, 4x8, 8x4, 4x4.

Vecteurs de mouvement Chaque partition ou partition de sous-macrobloc à l'intérieur d'un macrobloc codé en inter est prédite à partir d'une zone de même taille d'une image de référence. La différence entre les deux zones (le vecteur de mouvement) a une résolution au quart d'échantillon pour la luminance et au huitième d'échantillon pour la chrominance. Les échantillons de luminance et de chrominance aux positions sous-échantillonnées n'existent pas dans l'image de référence. Il est donc nécessaire de les créer par interpolation à partir d'échantillons du voisinage déjà codés.

Prédiction des vecteurs de mouvement Encoder un vecteur de mouvement pour chaque partition peut coûter un nombre de bits significatifs, surtout si des partitions de petites tailles sont choisies. Les vecteurs de mouvement des partitions voisines sont souvent hautement corrélés. Ainsi, chaque vecteur de mouvement est prédit à partir des vecteurs des partitions voisines déjà codées. Un vecteur prédit est formé à partir des vecteurs de mouvement précédemment calculés et la différence entre le vecteur courant et le vecteur prédit est codée et transmise. La méthode de formation de la prédiction dépend de la taille de la partition de la compensation de mouvement et de la disponibilité des vecteurs voisins.

Comme nous venons de le voir, les zones de prédiction contiennent très peu d'énergie après l'élimination de la redondance qu'elles contiennent. Ainsi, il n'est pas intéressant d'insérer la marque dans ces zones de prédiction puisqu'elle sera forcement endommagée ou complètement effacée.

La prédiction Intra

Ce mode de prédiction consiste à estimer les échantillons d'un macrobloc en utilisant exclusivement l'information contenue dans les blocs contigus appartenant au passé spatial de l'image courante. Ces blocs de référence doivent déjà avoir été encodés puis décodés par le codeur (de manière à retrouver des résultats identiques au codeur et au décodeur). La norme H.264/AVC propose deux types de traitement

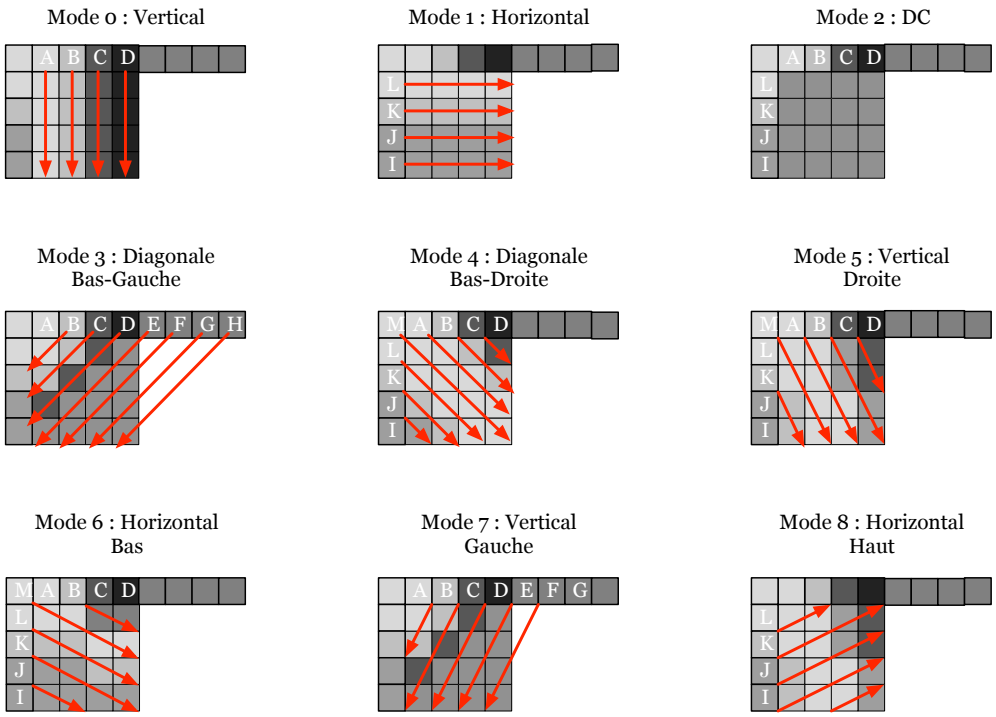


FIGURE 6.5 – Les neuf modes de la prédiction Intra4x4.

Intra pour prédire le signal de luminance. Le premier mode est appelé Intra4x4 et le second Intra16x16. Dans le type Intra4x4, le macrobloc est divisé en seize blocs de 4x4 pixels et chaque bloc est encodé individuellement. Neuf modèles de prédiction sont fournis par la norme et l'objectif du codeur est de sélectionner le mode le plus adapté au bloc courant. Ces neuf modes sont représentés sur la Fig.6.5. Nous pouvons constater que huit de ces modes caractérisent des modèles de prédiction directionnels. Chaque valeur prédite est déterminée à partir d'une combinaison linéaire entre les échantillons connus. Seul le mode DC n'est pas associé à une direction spécifique et consiste juste à calculer la moyenne des pixels contenus dans les deux blocs contigus du bloc courant.

Lorsque le type Intra16x16 est utilisé, seul un mode de prédiction est appliqué pour encoder l'ensemble du macrobloc. Quatre modèles de prédiction sont définis dans la norme et l'objectif du codeur est de choisir le mode le plus adapté au macrobloc courant. Ces quatre modes sont représentés sur la Fig.6.6. Nous pouvons noter que trois de ces modes sont associés à des directions spécifiques (verticales, horizontales et planes). Le dernier mode (DC) correspond à un calcul de moyenne. La prédiction Intra16x16 s'avère très efficace pour coder les régions où la luminance varie très peu car l'estimation s'applique au macrobloc complet. Pour le codage

des macroblocs de chrominance U et V, seule la prédiction Intra8x8 est utilisée. Ce schéma est suffisant car les variations des signaux de chrominance sont très faibles. Le mode Intra8x8 permet d'estimer les blocs 8x8 pixels d'un macrobloc de chrominance en proposant quatre modèles de prédiction : DC, verticale, horizontale et plane. Les types Intra8x8 et Intra16x16 sont donc identiques à un facteur d'échelle près. Il faut noter que les frames I, pour lesquelles uniquement une prédiction intra leur est appliquée, sont privilégiées pour l'insertion de l'information. Ceci est expliqué par le fait que l'énergie contenue dans les résidus de ce type de frames est beaucoup plus importante que le reste des frames vidéo. Ainsi, les macrosblocs des frames I absorberent mieux les distorsions et donc il est possible d'insérer une grande quantité d'information sans que cela soit perceptible. Cependant, il nous est impossible de dire à ce stade des travaux si les blocs prédits avec le mode Intra4x4 sont plus privilégiés que ceux prédits avec le mode Intra16x16 ou alors les deux cas sont équivalents. Ces derniers concernent les zones homogènes, donc les coefficients basses-moyennes fréquences sont plus important et donc nous auront un support presque parfait pour l'insertion de la marque. Malheureusement, les zones homogènes présentent une redondance spatiale importante et donc ça nous ramènerait au même problème que celui rencontré dans la prédiction Inter. De plus, marquer des blocs de zones homogènes risque d'être plus visible que marquer les blocs des zones texturées. D'un autre côté, les blocs prédits avec le mode Intra4x4 appartiennent aux zones texturées donc leurs tatouage sera moins visible. Par contre, le nombre de coefficients basse-moyennes fréquences seront très peu présents et donc le support pour insérer la marque sera très réduit et l'utilisation de l'étalement risque de poser des problèmes. Enfin, il peut ête intéressant de marquer aussi les blocs de la composante chrominance, bien que ce ne soit pas très courant dans la communauté. Notons que l'oeil humain est très peu sensible à la composante chrominance, ceci peut nous permettre d'insérer au moins une partie de la marque dans ces composantes.

Le codage par transformée

Dans le même contexte que les normes précédentes, la phase de transformation-quantification est appliquée dans le but de coder le signal d'erreur de prédiction. La tâche de la transformée consiste à réduire les redondances spatiales du signal d'erreur. Le H.264/AVC dispose d'une transformée basée sur des entiers. La taille de la matrice de transformation est généralement composer de 4x4 éléments, mais

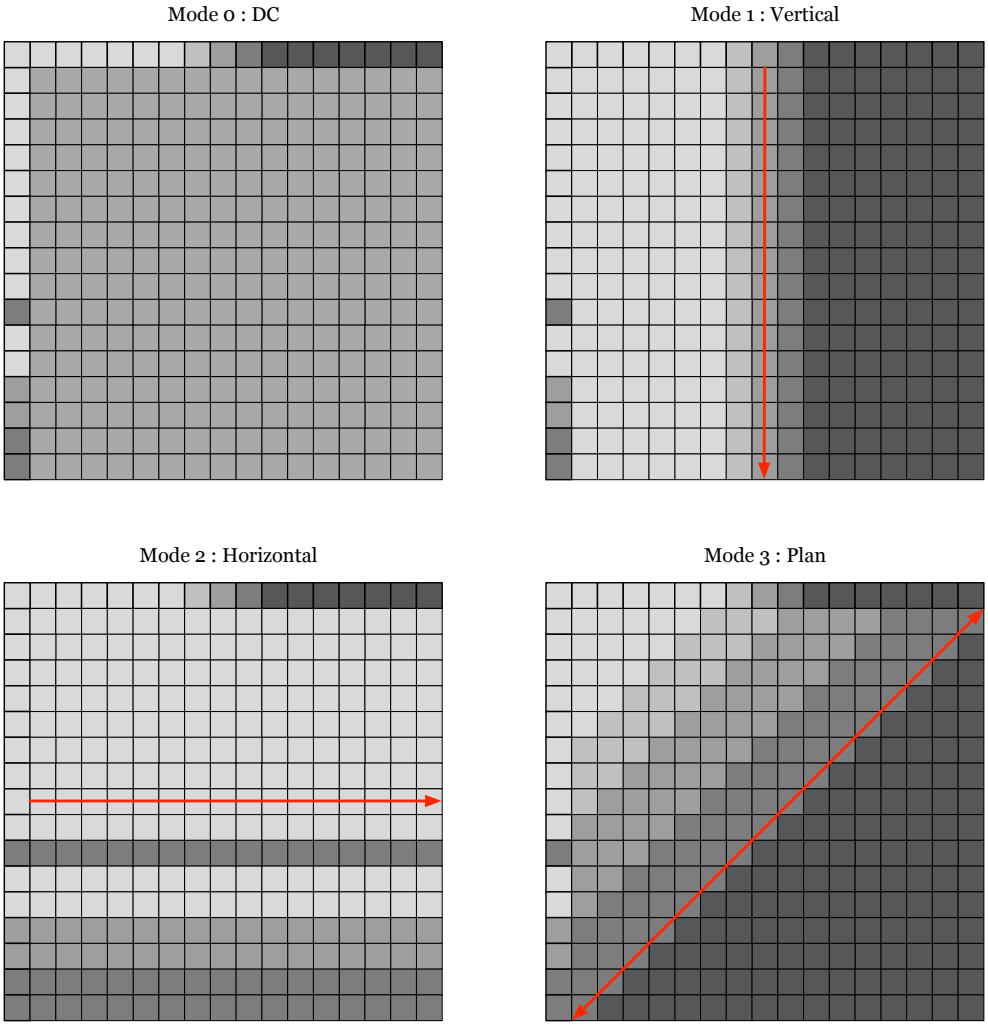


FIGURE 6.6 – Les quatre modes de la prédiction Intra16x16.

peut être réduite à 2x2 composantes pour le codage de certaines informations de chrominance. La diminution de la taille de la fenêtre d'analyse permet à l'encodeur de mieux adapter le codage de l'erreur de prédiction aux frontières des objets mouvants. En effet, la taille du bloc est similaire aux dimensions de la plus petite zones d'analyse de l'estimation Inter ou Intra (4x4 pixels) et la transformée s'ajuste donc mieux aux erreurs de prédiction locales. Il existe trois types de transformées. Le premier est appliqué à tous les échantillons d'erreurs à la fois pour le signal de luminance Y mais aussi pour les composantes de chrominance U et V, quelque soit le mode de prédiction utilisé (Inter ou Intra). Cette matrice de transformation H1 est composée de 4x4 éléments. Sa structure est donnée par ce qui suit :

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 1 & -1 & -2 \\ 1 & -1 & -1 & 1 \\ 1 & -2 & 2 & -1 \end{bmatrix} \tag{6.1}$$

Si le macrobloc est prédit en utilisant le mode Intra16x16, la seconde transformée est appliquée en plus de la première. Cette dernière convertit les seize coefficients DC des blocs transformés d'un macrobloc de luminance. Elle correspond à une transformée de Hadamard dont la taille s'élève à 4x4 composantes. Une représentation de la matrice est donnée par :

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix} \tag{6.2}$$

Le troisième type se rapporte aussi à une transformée de Hadamard mais de taille 2x2. Elle est utilisée pour le codage des quatre coefficients DC contenus dans un macrobloc de chrominance. Sa matrice est donnée par :

$$\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \tag{6.3}$$

L'opération de transformation dans H.264/AVC se traduit par l'équation

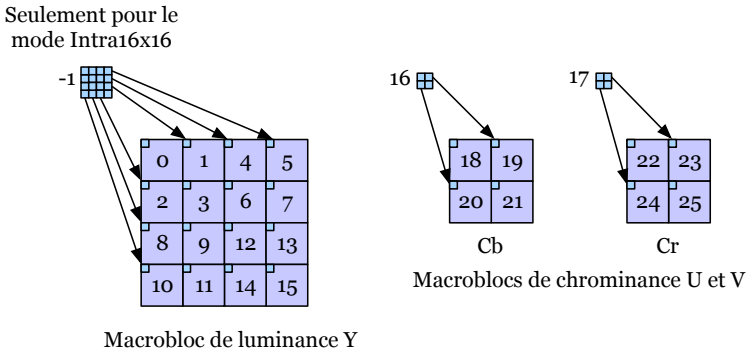


FIGURE 6.7 – Ordre de transmission de tous les coefficients d'un macrobloc.

$$Y = H_i \times X \times H_i^T, \tag{6.4}$$

où Y est la matrice transformée, X est le signal d'entrée et H_i peut représenter H_1, H_2 ou H_3. L'ordre de transmission de tous les coefficients est donné sur Fig.6.7. Si le macrobloc est prédit en utilisant le mode Intra16x16, le bloc muni de l'index -1 est diffusé en premier. Ce paquet contient les coefficients DC de tous les blocs de luminance. Tous les ensembles indicés de 0 à 25 sont ensuite transmis. Les éléments numérotés de 0 à 15 correspondent à tous les coefficients AC de la luminance. Les blocs 16 et 17 représentent les composantes DC de chaque signal de chrominance. Enfin, les valeurs indexées de 18 à 25 se rapportent aux coefficients AC de la chrominance.

Comparées à la DCT, les matrices de transformation du H.264/AVC sont composées seulement de nombres entiers dans un intervalle compris entre -2 et $+2$. Ce principe permet de calculer la transformée et son inverse sur seize bits en utilisant seulement des opérations de décalages, des additions et des soustractions. Dans le cas d'une projection de Hadamard, seules l'addition et la soustraction sont nécessaires. De plus, les disparités liées aux approximations du calcul flottant sont complètement évitées grâce à l'utilisation exclusive d'opérations sur des entiers. Tous les coefficients sont ensuite quantifiés par le biais d'un quantificateur scalaire. La taille du pas de quantification est choisie par un paramètre QP qui supporte cinquante deux valeurs possibles. La taille du pas double lorsque la variable QP est incrémentée de 6. Une augmentation de QP de 1 entraine un accroissement du débit des données d'environ $12,5\%$.

Il est clair que cette étape transformation-quantification est la plus critique lors-

qu'un tatouage est appliqué à la version bande de base de la vidéo. Ceci est dû au fait qu'elle représente l'étape de perte d'informations. Afin de contourner ce problème, il est nécessaire d'insérer la marque dans les zones les plus pertinentes, qui ne seront jamais effacé, et où l'on prend le moins de risque en les modifiant. Les coefficients basses-moyennes fréquences obtenus après transformations sont des cibles de choix pour insérer la marque. Cependant, il serait intéressant de mener des expérimentations sur le marquage des coefficients DC des blocs appartenant à des zones moyennement texturées à texturées afin de vérifier l'impact visuel qu'aura un tatouage des coefficients DC.

Le codage entropique

La procédure d'encodage entropique introduit les propriétés sémantiques et synthaxiques dans le flux vidéo compressé. Le H.264/AVC propose deux méthodes alternatives de codage entropique : une technique de faible complexité basée sur l'usage de contextes adaptatifs contenant des codes de longueurs variables, nommée CAVLC (Context-based Adaptive Variable Length Coding) [83], et un algorithme plus coûteux basé sur un codage arithmétique reposant sur des tables évolutives, le CABAC (Context-based Adaptive Binary Arithmetic Coding) [84]. Les deux méthodes représentent des améliorations majeures en termes d'efficacité de compression en comparaison avec les techniques de codage statistique traditionnelles. Dans les anciens standards, l'encodage de chaque élément de syntaxe était basé sur des tables VLC fixes (une distribution de probabilité était associée à chaque élément). Cependant, des études pratiques ont rapidement démontré que les signaux étaient rarement stationnaires et que l'utilisation de tables adaptatives (contextuelles) était plus efficace pour compresser les données. Des modèles contextuels ont donc été intégrés dans le processus d'encodage entropique. Pour ce travail, nous utiliseront exclusivement la méthode CAVLC pour deux raisons : elle fournit un codage efficace de faible complexité et elle est plus adaptée aux applications liées aux télécommunications.
Dans le codage CAVLC, deux techniques de compression sont utilisées :

- La première, basée sur un codage Exponential-Golomb (noté Exp-Golomb dans la suite) [85], se charge de compresser tous les paramètres de codage (type de macrobloc, pas de quantification, vecteurs de mouvement, etc) à l'exception des résidus de prédiction.
- La deuxième s'occupe de compresser les résidus de prédiction, plus compliquée,

mais permettant de compresser les données de manière optimale.

Le codage Exp-Golomb Dans ce type de codage, tous les symboles d'entrée sont représentés par des entiers. Ils sont ensuite convertis en codes VLC par l'intermédiaire de la table de correspondance (voir la Tab.6.1). Certains symboles peuvent correspondre à des valeurs positives ou nulles, d'autres à des nombres signés. A titre d'exemple, le mode de prédiction Intra16x16 d'un macrobloc est défini par une valeur entière allant de 0 à 3 (pour les quatre modes proposés par la norme). A l'inverse, les vecteurs de mouvements sont des valeurs entières positives, négatives ou nulles. Nous pouvons constater que les codes Exp-Golomb sont composés d'un préfixe et d'un suffixe. Le préfixe représente la première portion du code, il contient un ensemble de zéros et se termine par 1. Le nombre de zéros contenus dans le préfixe spécifie la taille du suffixe.

Lors de la réception, le décodeur récupère le préfixe et déduit la taille binaire du

Valeur positive	Valeur signée	Mot de code
0	0	1
1	+1	010
2	−1	011
3	+2	00100
4	−2	00101
5	+3	00110
6	−3	00111
7	+4	0001000
8	−4	0001001
9	+5	0001010
10	−5	0001011
...	...	...

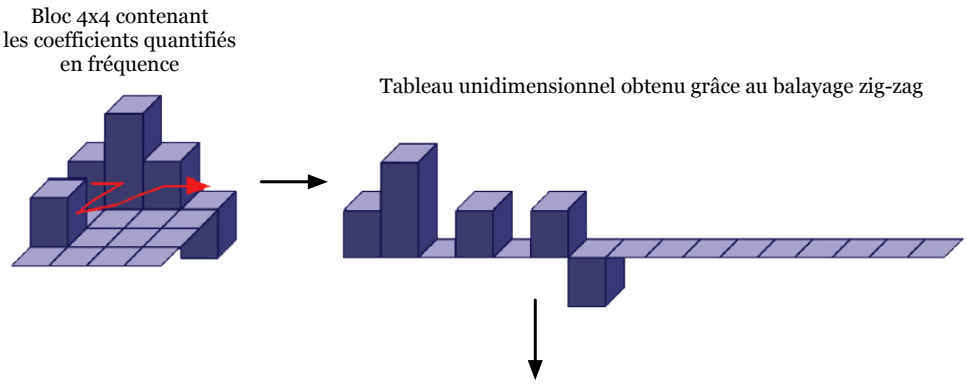
TABLE 6.1 – Table de correspondance du codage Exp-Golomb.

suffixe. La table représentée sur Tab.6.1 nous montre également que les codes les plus courts sont attribués aux symboles les plus faibles (en valeur absolue). La distribution de probabilité d'un symbole doit donc être concentrée autour de 0. Cependant, tous les symboles ne respectent pas directement cette distribution. C'est le cas des vecteurs de mouvement ou du pas de quantification par exemple. L'encodeur doit alors procéder à une étape de prédiction ou de translation. Concernant les vecteurs de mouvements, ces derniers sont prédits dans une phase préliminaire en utilisant

les vecteurs de mouvement des blocs situés au-dessus et à gauche du bloc courant. Seul l'offset résiduel (résultant de la différence entre le vecteur exact et sa prédiction) est encodé en Exp-Golomb, puis transmis. Lors de l'opération de codage du pas de quantification moyen d'une image, les valeurs possibles varient entre 0 et 51. L'encodeur soustrait donc la valeur choisie de 26 pour ramener l'ensemble des valeurs dans l'intervalle $[-26; +25]$. La valeur résultante est ensuite codée à partir de la table Tab.6.1.

Il est évident qu'il n'est pas intéressant de tatouer les valeurs codées avec Exp-Golomb. La raison pour le H.264 est que les entités concernées sont trop importantes et toute modification entraînerait beaucoup de distorsions qui impacteront la qualité visuelle de la vidéo.

Le codage des résidus Nous allons détailler ce point parce que les levels sont capables d'absorber les distorsions dues au marquage et donc présentent un intérêt particulier pour l'insertion de l'information. Ceci est d'autant plus vrai lorsque les levels dépassent un certain seuil.
Dans le codage CAVLC, les résidus de prédiction (transformés et quantifiés) sont compressés de manière indépendante en suivant une procédure spécifique. Cette méthode est résumée sur Fig.6.8. Les seize coefficients du bloc transformé sont d'abord répartis dans un tableau unidimensionnel en respectant un schéma de balayage spécifique (zig-zag scan). Cette technique permet notamment de regrouper les coefficients quantifiés faibles ou nuls à la fin du tableau. Après cette opération, les zéros sont statistiquement rassemblés dans les hautes fréquences et une occurrence prédominante de niveaux égaux à +/- 1 (notés T1s) sont situés à la fin du tableau. Le nombre de coefficients non-nuls (TotalCoeffs) et le nombre de T1s (TrailingOnes) sont encodés en premier lieu. Ces deux paramètres sont combinés dans un simple mot de code (CoeffToken), extrait d'une des tables VLC prédéfinies. La sélection de la table VLC dépend du nombre de coefficients non-nuls contenues dans les blocs voisins. Dans une deuxième étape, le signe et l'amplitude de chaque coefficient non-nul sont encodés en parcourant le tableau dans le sens inverse (en partant des hautes fréquences pour aboutir aux basses fréquences). Pour l'encodage des T1s, seul le signe est nécessaire qui est représenté par un seul bit. Concernant les autres coefficients non-nuls (Levels), la procédure est plus complexe. Cette technique considère que l'amplitude des coefficients augmente lorsque la fréquence diminue. De manière sim-



	Codage CAVLC
Préambule	TotalCoeffs = 5 TrailingOnes = 3
Eléments de syntaxe parcourus dans le sens inverse	Signe T1s : -1, +1, +1 Levels : +2, +1 TotalZeros = 2 RunBefores : 0, 1, 1

FIGURE 6.8 – Codage des résidus de prédiction d'un bloc 4x4 avec la méthode CAVLC.

plifiée, la table VLC, utilisée pour encoder l'amplitude et le signe d'un coefficient, est construite en fonction du coefficient précédent. Le nombre de coefficients nuls précédant le dernier coefficient non-nul (TotalZeros) est encodé dans une troisième étape. Son encodage dépend des tables VLC prédéfinies. La table appropriée est choisie en fonction de la valeur de TotalCoeffs. Le paramètre TotalZeros est suivi des RunBefores qui indiquent le nombre de zéros consécutifs situés entre chaque coefficient non-nul. Chaque RunBefore est codé séparément en utilisant des tables prédéfinies. La table adaptée dépend du nombre de coefficients nuls restant à encoder.

Level Prefix	Mot de code
0	1
1	01
2	001
3	0001
4	00001
5	000001
6	0000001
7	00000001
8	000000001
9	0000000001
10	00000000001
11	000000000001
12	0000000000001
13	00000000000001
14	000000000000001
15	0000000000000001

TABLE 6.2 – Table VLC utilisée pour l'encodage du Level Prefix.

Codage des Levels Les Levels (amplitudes et signes des coefficients non-nuls restants) sont également encodés en respectant un ordre de balayage inverse. L'amplitude et le signe de chaque coefficient sont incorporés dans le paramètre Code en suivant la procédure définie par les deux équations suivantes :

$$\begin{aligned} Code &= (amplitude - 1) \ll 1 \\ Code &= Code \cup Signe, \end{aligned} \tag{6.5}$$

Où $\ll$ réalise un décalage binaire de 1 bit vers la gauche et $\cup$ représente l'opérateur OR (ou logique). Le signe est codé sur le bit de poids faible et l'amplitude sur les bits de poids forts.
Le paramètre Code est ensuite décomposé en deux éléments de syntaxe : LevelPrefix et LevelSuffix. Les valeurs numériques de ces éléments sont déterminées à partir de ce qui suit

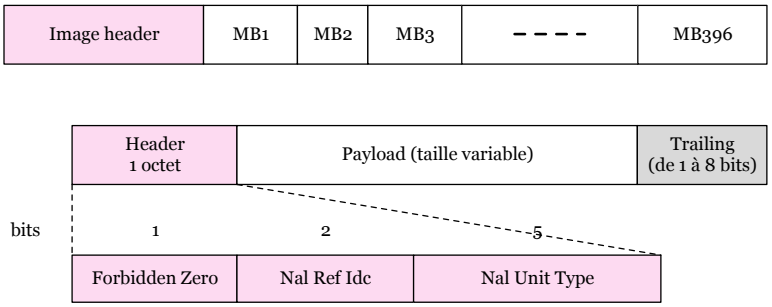
$$\begin{aligned} LevelPrefix &= Code \gg Shift \\ LevelSuffix &= Code(Shift), \end{aligned} \tag{6.6}$$

tel que réalise un décalage de 1 bit vers la droite et Code(Shift) représentent la valeur correspondant aux Shift bits de poids faibles du Code. Finalement, l'élément de syntaxe *LevelPrefix* est encodé à l'aide de la table VLC exposée sur la Tab.6.2 et les Shift bits de *LevelSuffix* sont ajoutés à la suite. Le principe d'adaptabilité dans la phase d'encodage des Levels repose essentiellement sur la variable Shift. Cette valeur peut évoluer au cours de la phase d'encodage des Levels. Pour le premier coefficient, elle est initialisée à 0. Pour les cycles suivants, elle est incrémentée de 1 si l'amplitude du coefficient courant (déjà encodé) est supérieure au seuil associé à la valeur courante de Shift, définie dans la table de la Tab.6.3. Cette technique permet à l'encodeur de s'ajuster au mieux à la dynamique des coefficients restant à encoder. Sachant que l'évolution des amplitudes a tendance à croître lorsque la fréquence diminue, cette technique s'avère très efficace dans cette situation.

Shift	Valeur du seuil
0	0
1	3
2	6
3	12
4	24
5	48
6	∞

TABLE 6.3 – Valeurs des seuils associés à la variable Shift.

L'opération de décodage est composée de plusieurs étapes. Dans un premier temps, le décodeur commence par récupérer *LevelPrefix*. Il extrait ensuite l'élément *LevelSuffix* du flux en se basant sur la valeur courante du Shift. En inversant les équa-



FIGUI ... CIF.

FIGURE 6.10 – Format d'un paquet NAL.

tions précédentes utilisées pour calculer les valeurs Code *LevelSuffix* et *LevelPrefix*, le décodeur arrive à retrouver l'amplitude et le signe associés à chaque coefficient. Par ailleurs, il met à jour la valeur de Shift à chaque itération.

6.2.2 La couche réseau (NAL : Netework abstraction Layer)

Dans les parties précédentes, nous avons constaté que la couche VCL fournit les outils nécessaires pour compresser une image sur un nombre restreint de bits. Les macroblocs d'une image sont encodés successivement et le flux généré par la couche VCL correspond donc à une succession de séquences binaires caractérisant chaque macrobloc encodé. Le format du flux compressé pour une image CIF (Common Intermediate Format, 352 x288 pixels) est illustré sur la Fig.6.9.

Un en-tête (Image header) est inséré au début du flux et définit la valeur des paramètres généraux de l'image. En revanche, une vidéo compressée ne se limite pas à une succession d'images encodées. D'autres informations sont nécessaires pour assurer la compatibilité entre l'encodeur et le décodeur. Par conséquent, la couche VCL génère des flux complémentaires qui contiennent les paramètres additionnels de la vidéo. Ces flux de données sont insérés entre les images lors du stockage ou de la transmission. A titre d'exemple, ces flux peuvent représenter les paramètres importants s'appliquant à une large séquence d'images (SSP ou Sequence Parameter Set) ou définissant les caractéristiques liées à l'affichage des images (SEI ou Supplemental Enhancement Information). Contrairement à la couche VCL, la couche NAL convertit la structure VCL dans un format approprié aux systèmes de communication actuels. A un niveau général, la vidéo encodée est répartie dans des paquets NAL qui contiennent un nombre entier d'octets. Les informations utiles (flux VCL) sont encapsulées dans un champ de contenu (Payload). Un en-tête (Header) de 1 octet est ajouté au début de chaque paquet et précise le type de données transpor-

tées. Un champs de bourrage (Trailing bits) est ajouté à la fin de chaque paquet et permet de ramener la taille du paquet à un nombre entier d'octets. Le format des paquets NAL est illustré sur la Fig.6.10. Les champs d'information composants le paquet sont spécifiés ci-dessous :

- Les champs ForbiddenZero de 1 bit doiventt toujours être égaux à 0.
- Les 2 bits du champ NalRefIdc spécifient le degré d'importance des informations contenues dans le paquet. Plus précisément, ce paramètre indique si l'image encodée du paquet correspond à une image de référence. Plus cette valeur est grande (tend vers 3), plus l'image a d'influence sur la qualité vidéo générale.
- Les 5 bits du champ NalUnitType spécifies la nature des données transportées dans le paquet. Il existe différentes catégories d'informations (image encodée, SEI, SSP, etc) et le NalUnitType définit ce type. Le champ NalUnitType contenu dans l'en-tête des NALs permet de détecter les images Intra. Plus précisément, lorsque NalUnitType est égal à 5 en décimal (voir norme), le NAL contient une image Intra comprimée et les blocs transformés de cette image seront marqués.
- La payload contient les données utiles à transmettre : elle représente les informations fournies par la couche VCL. Ce champs peut contenir un nombre variable de bits car chaque image encodée a une taille variable.
- Le champ Trailing représente des données de remplissage. Son bit de poids fort vaut 1 et les bits suivant sont fixés à 0. Ces bits permettent d'ajuster la taille du paquet pour obtenir un nombre entier d'octets.

Afin de mettre au point un filtre qui permettrait de garder uniquement les paquets correspondants à une image de référence I dans un flux binaire H.264, nous utiliserons le Header pour repérer et isoler les images en question. Nous nous intéresserons donc aux 5 bits que contient le NalUnitType, puisque nous garderons seulement les paquets dans la séquence NalUnitType correspondrants à une image de référence I. De plus, il est possible que nous analysions d'abords les 2 bits de NalRefIdc pour faire une première sélection afin de minimiser l'erreur et surtout réduire le temps d'analyse en comparant les différentes séquences à une table contenant 4 éléments au lieu d'une table de 32 éléments à chaque fois.

6.3 Procédé d'insertion du Fingerprint

Ayant décrit les principaux éléments qui entrent dans la conception d'un flux compressé H.264, nous proposons une technique de traçage résultante d'un compromis qui tient compte de ces différents éléments et surtout des contraintes qu'ils imposent au système. Le sécurisation du flux binaire vidéo se fait dans un domaine compressé tel qu'il est décrit sur Fig.6.1.
Un tatouage dans un domaine compressé peut présenter des avantages mais aussi des inconvénients. Dans ce cas, l'attaquant est obligé de disperser la puissance de son attaque sur toute la vidéo puisque les composantes marquées pourront se retrouver dispersées sur une grande partie de la vidéo après décompression. Dans le cas où cet attaquant serait tenté par un deuxième ré-encodage, il prendra le risque de perdre énormément en terme de compression puisque les informations non-pertinentes dans la vidéo ont déjà été éliminées par la première compression (une deuxième compression qui détériorera la marque ne gardera jamais la même qualité visuelle qu'en premier). D'un autre côté, un flux compressé présente très peu d'endroits où il est possible de tatouer. En fait, ce flux est un ″concentré″ d'informations sur les séquences vidéo et une simple modification pourrait avoir des effets très visibles sur la vidéo. Donc, il est très important que toutes les zones vérifiant les conditions de tatouage soient identifiées, au préalable, pour définir avec précision les paramètres d'insertion de la marque qui permettront le meilleur compromis entre indivisibilité et robustesse.

En résumé, les facteurs qui ont orienté les choix du marquage des flux compressés sont les suivants :

1. Le H.264 définit un standard de compression avec perte, il est très optimisé pour produire un flux compressé très robuste aux dégradations, essentiellement dues au canal de transmission, avec un débit binaire relativement bas.
2. Toutes les zones redondantes et toutes les informations non-pertinentes sont automatiquement éliminées du flux compressé.
3. Les éléments constituants le flux H.264 ont un ordre d'importance et un impact visuel sur la vidéo très différents.
4. L'énergie présente dans les frames change d'une scène à l'autre, donc la taille du support de tatouage ne sera pas fixe pour toutes les vidéo.
5. Le codage entropique tient compte des statistiques du signal, tout changement

dans le signal risque donc d'augmenter la taille du fichier compressé.

6. La taille du flux vidéo doit rester constante avant et après insertion de la marque. Tout changement dans la taille du contenu transmis risque de compromettre la bonne lecture du fichier de la part du client.

6.3.1 Domaine d'insertion de l'information

Comme dans le cas des systèmes de tatouage résistant à la compression JPEG (pour lesquels la marque est insérée dans les coefficients DCT qui ne sont pas éliminés par l'encodeur), le domaine qui sera ciblé par l'insertion de la marque est, idéalement, celui qui est épargné par les pertes de compression. Dans notre cas, la norme H.264 élimine surtout les coefficients AC contenant le moins d'énergie et correspondant aux hautes fréquences. Donc, l'insertion se fait dans les coefficients basses-moyennes fréquences de la transformation basée sur des entiers [86] (décrite dans la section précédente). Si par exemple l'insertion se fera dans un autre domaine et avec une transformation autre que celle utilisée par le standard H.264, il y a de fortes chances que les informations insérées, ou du moins une partie, seront complètement perdues ou fragilisées face aux manipulations licites. Il faut noter aussi que les coefficients basses-moyennes fréquences correspondent aux coefficients ayant les plus grandes valeurs dans le bloc transformé, si ces valeurs sont assez élevées l'impact visuel qu'aura le tatouage sur la vidéo sera très limité.

6.3.2 Localisation de l'information dissimulée

Dans le cas du H.264, il est préférable d'insérer la marque uniquement dans les levels des coefficients des images Intra. Pourquoi les images Intra ? Après compression se sont les frames dont les résidus contiennent le plus d'énergies. Ils seront donc les mieux préservés pour servir de séquences de références pour la prédiction des autres frames. De plus, vu leurs importances, un attaquant n'aura pas intérêt à trop distordre ces frames pour enlever la marque au risque de rendre la vidéo inutilisable. D'un autre côté, nous choisissons les levels des coefficients non-nuls parce qu'ils ont généralement des valeurs assez importantes qui permettent au système de marquage de préserver la qualité visuelle de la vidéo.

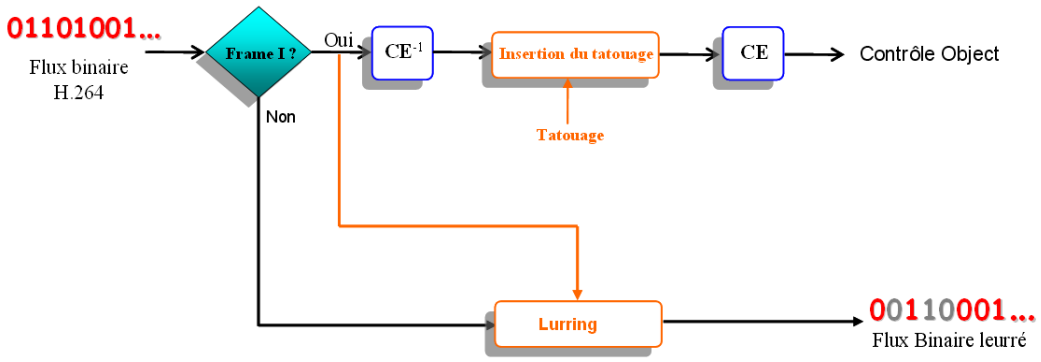


FIGURE 6.11 – Schéma d'insertion d'un Fingerprint dans un flux binaire H.264.

6.3.3 Schéma d'insertion

La procédure d'insertion de la marque est résumée à l'aide du schéma donné sur Fig.6.11. Nous commençons d'abords par sélectionner les Frames I et ceci en gardant uniquement les paquets du flux H.264 ayant un NalUnitType correspondant à une séquence de référence I. Un codage entropie inverse est ensuite appliqué au paquet correspondant afin de reconstruire les coefficients issus de la transformation lors de la compression. Afin d'insérer une marque qui soit parfaitement invisible, il est préférable de tatouer que les coefficients ayant une valeur dépassant un seuil fixé selon la puissance de la marque avec laquelle on voudrait faire l'insertion. Cependant, dans nos experiences nous n'avons pas tout le temps respecté cette condition à cause de contraintes d'ordre pratique.
Le problème est que toute modification dans le flux binaire même si elle est minime risque de modifier la taille de notre contrôle object. Ceci risque de compromettre la phase de décompression de la vidéo. La solution serait d'insérer la marque en modifiant les suffixes en gardant évidement la même taille pour ces suffixes.

6.3.4 Résultats préliminaires

Nous avons mené un certain nombre d'expériences, où nous avons inséré le tatouage au niveau des suffixes d'une vidéo sous format H.264. Nous avons pu réccupérer intégralement la marque du flux binaire compressé sans erreurs. Cependant, l'opération de décompression-recompression élimine une grande partie du tatouage et le rend quasi inutilisable, puisque pour 7 bits insérés dans les composantes luminance, nous n'avons pu réccupérer que 3 bits d'information. Alors que pour les composantes chrominance, nous avons pu réccupérer 18 bits d'information sur 28 insérés.

Pour rendre le tatouage plus résistant, la solution a été d'insérer la marque au niveau des suffixes mais la marque doit être réccupérée au niveau des coeffcients DCT ou des Pixels.
D'un autre côté, nous avons constaté que l'impacte visuel du tatouage est très limité et la taille du flux binaire reste la même qu'avant insertion. Ainsi, pour une séquence de la vidéo foreman qui a une durée de 5 secondes, le PSNR avant et après tatouage est resté égal à $45,4dB$ pour la composante luminance, alors que pour les composantes chrominances le PSNR reste égal, respectivement, à $48,8$ pour la composante U et $51,3$ pour la composante V. Ce bon niveau d'imperceptibilité est expliqué par le fait que l'effet de la marque est réduit grace à l'utilisation de la QIM (golablement, 50% des coefficients hôtes ne sont pas modifiés). De plus, les traitements effectués par le standard H.264 au décodage permettent d'étaler les modifications dues à la marque sur plusieurs séquences et donc moins perceptibles à l'utilisateur.

6.4 Conclusion

Dans ce chapitre, nous avons proposé un système de protection des contenus vidéo en utilisant un schéma d'insertion informé. Ainsi en se basant sur la bibliographie et sur les particularités du flux H.264, nous avons proposé un schéma d'insertion pour permettre la protection d'un flux vidéo compressé MPEG4-AVC/H.264. Ce schéma d'insertion tiens compte des contraintes liées au standard H.264 et l'impacte perceptuel après décompression. Les résultats préliminaire obtenus à ce stade des travaux sont très encourageants puisque même s'il reste des améliorations concernant la résistance face à la recompression, l'insertion se fait de manière imperceptible et la réccupération de la marque se fait sans erreur sur le flux compressé.
Il est important de noter qu'il existe des extensions possibles pour le schéma proposé :

1. Suppression des étapes de codage entropique inverse et du codage entropique : après avoir détecté une frame I, nous allons localiser les levels qui dépasseront le seuil que nous fixons dans le paquet de la frame. La séquence binaire correspondante sera directement remplacée par une séquence de levels correspondante à l'information qu'on voudrait transmettre. Ceci est facilité par le fait que le codage des levels est indépendant des autres paramètres.
2. Utilisation d'une autre technique de tatouage informé : Il est possible que l'utilisation d'un code correcteur d'erreur combiné à la QIM s'avère plus résistante

face à une ré-compression H.264 que la QIM basique. Cette extension possible est motivée par la nature de l'attaque par re-compression.

Chapitre 7

Appendix B
Spread Transform contre l'attaque TFA

Considérons un signal hôte vidéo représenté par le vecteur : $\mathbf{s} = [\mathbf{s}[0], ..., \mathbf{s}[M-1]]^T, M \in \mathcal{N}^*$. Le ST procède à un étalement d'une information donnée sur τ échantillons du signal hôte, où τ est un entier non nul représentant facteur d'étalement. Ainsi, il décompose le vecteur signal hôte $\mathbf{s}$ en M/τ vecteurs, i.e., $\mathbf{s} = [\mathbf{s}_0, \mathbf{s}_1, ..., \mathbf{s}_{M/\tau}]$. De même, le vecteur projection : $\mathbf{t} = [\mathbf{t}[0], ..., \mathbf{t}[M-1]]$ est décomposé en un ensemble de vecteurs de taille τ : $\mathbf{t} = \mathbf{t}_0, \mathbf{t}_1, ..., \mathbf{t}_{M/\tau}$, les vecteurs $\mathbf{t}_i$ vérifie la condition normalisation $< \mathbf{t}_i, \mathbf{t}_i >= 1$, où $<,>$:représente le produit scalaire entre deux vecteurs.
Notons que la décomposition du signal hôte et du signal marqué en plusieurs vecteurs de taille τ est appelée parfois ″framing″, nous désignons alors ces vecteur par frame. Ainsi, La l^{ieme} frame du signal marqué est donnée par :

$$\mathbf{x}_l = \mathbf{s}_l + (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_l$$

tel que $\mathbf{s}^{ST}[l]$ représente la transformation du vecteur signal hôte $\mathbf{s}$, donnée par la formule suivante : $\mathbf{s}^{ST}[l] = \sum_{i=\tau l}^{\tau l+\tau-1} \mathbf{s}[i] \cdot \mathbf{t}[i] = \sum_{i=0}^{\tau} \mathbf{s}_l[i] \cdot \mathbf{t}_l[i]$ et $\mathbf{x}^{ST}[l]$ représente la version tatouée (dans le domaine transformé) de $\mathbf{s}^{ST}[l]$.
Le tatouage se fait séquence par séquence, si nous supposons que le signal marqué a été envoyé à travers un canal gaussien AWGN, tel que le signal bruit ajouté est donné par le vecteur : $\mathbf{v} = [\mathbf{v}[1], \mathbf{v}[2], ..., \mathbf{v}[M-1]]$, et le signal tatoué attaqué :

$\mathbf{y} = [\mathbf{y}[0], ..., \mathbf{y}[M-1]]$. Comme pour le signal hôte, $\mathbf{y}$ est formé de plusieurs vecteurs (frames) de taille τ : $\mathbf{y}_0, \mathbf{y}_1, ..., \mathbf{y}_{|[M/\tau]|}$, tel que :

$$\mathbf{y}_l = \mathbf{s}_l + (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_l + \mathbf{v}_l \tag{7.1}$$

Pour extraire l'information, il faut calculer la projection $\mathbf{y}^{ST}[l]$:

$$\mathbf{y}^{ST}[l] = \sum_{i=\tau l}^{\tau l+\tau-1} \mathbf{y}[i] \cdot \mathbf{t}[i] = \sum_{i=0}^{\tau-1} \mathbf{y}_l[i] \cdot \mathbf{t}_l[i] =< \mathbf{y}_l, \mathbf{t}_l >, l = 0, ..., |[\frac{M}{\tau}]|, \tag{7.2}$$

donc,

$$\begin{aligned} \mathbf{y}^{ST}[l] &= < \mathbf{y}_l, \mathbf{t}_l > =< \mathbf{s}_l + (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_l + \mathbf{v}_l, \mathbf{t}_l > \\ &= < \mathbf{s}_l, \mathbf{t}_l > + < (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_l, \mathbf{t}_l > + < \mathbf{v}_l, \mathbf{t}_l > \\ &= \underbrace{< \mathbf{s}_l, \mathbf{t}_l >}_{=\mathbf{s}^{ST}[l]} + \underbrace{(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])}_{=\mathbf{w}^{ST}[l]} \cdot \underbrace{< \mathbf{t}_l, \mathbf{t}_l >}_{=1} + \underbrace{< \mathbf{v}_l, \mathbf{t}_l >}_{\mathbf{v}^{ST}[l]}, \end{aligned} \tag{7.3}$$

alors :

$$\mathbf{y}^{ST}[l] = \mathbf{s}^{ST}[l] + \mathbf{w}^{ST}[l] + < \mathbf{v}_l, \mathbf{t}_l > . \tag{7.4}$$

Rappelons que la vidéo est une suite de séquences d'images qui sont composées elles mêmes de N pixels (échantillons) et qui forment un vecteur de pixels $\mathbf{s}_i$, où i est un entier donnant l'index de l'image (séquence) dans la vidéo hôte. Cette dernière est donc décomposé en un ensemble de vecteurs $\mathbf{s}_i$, i.e., la vidéo est aussi représentée par le vecteur : $\mathbf{s} = [\mathbf{s}_0, ..., \mathbf{s}_{|[M/N]}]|]$.

7.1 Cas d'un étalement sur une frame vidéo

Sachant qu'une frame ayant subit une attaque TFA est formulée comme suit (voir aussi l'exemple sur Fig.7.1),

$$\dot{\mathbf{y}} = \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{y}_u, \tag{7.5}$$

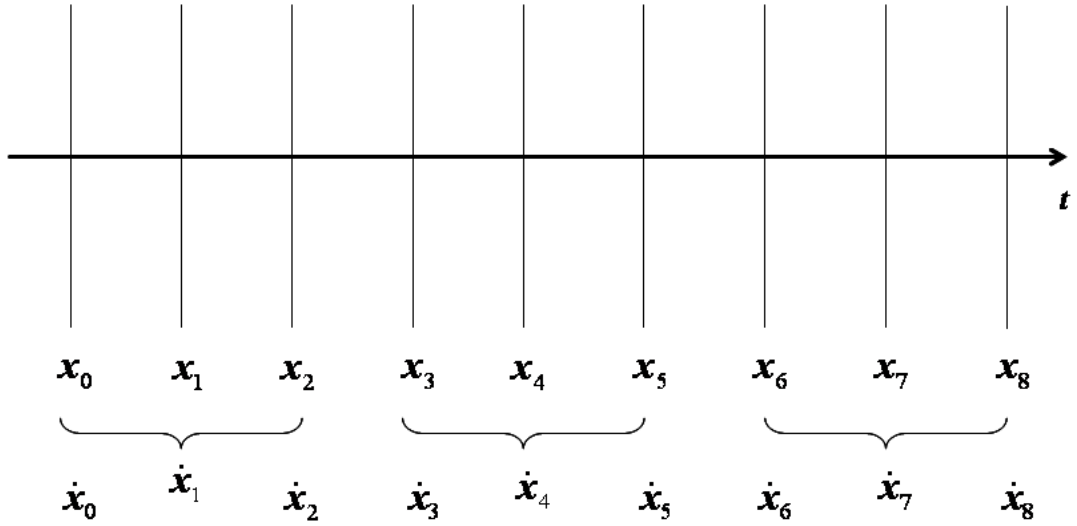


FIGURE 7.1 – Exemple d'une attaque TFA sur une suite de frames vidéo avec une fenêtre d'attaque ω égale à 3.

et d'après Eqn.7.1, le signal reçu après une attaque TFA est formulée comme suite :

$$\dot{\mathbf{y}}_l = \underbrace{\frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}\mathbf{s}_{l+u}}_{\dot{\mathbf{s}}_l} + \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}(\mathbf{x}^{ST}[l+u]-\mathbf{s}^{ST}[l+u])\cdot\mathbf{t}_{l+u} + \underbrace{\frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}\mathbf{v}_{l+u}}_{\dot{\mathbf{v}}_l} \quad (7.6)$$

Avant le décodage, une transformation est appliquée sur le signal reçu d'où,

$$\dot{\mathbf{y}}^{ST}[l] = \sum_{i=\tau l}^{\tau l-\tau-1}\dot{\mathbf{y}}[i]\cdot\mathbf{t}[i] = \sum_{i=Nl}^{Nl-N-1}\dot{\mathbf{y}}[i]\cdot\mathbf{t}[i] = \sum_{i=0}^{\tau-1}\dot{\mathbf{y}}_l[i]\cdot\mathbf{t}_l[i] = \sum_{i=0}^{N-1}\dot{\mathbf{y}}_l[i]\cdot\mathbf{t}_l[i] =< \dot{\mathbf{y}}_l, \mathbf{t}_l >, \quad (7.7)$$

alors, le signal reçu attaqué est formulé comme suit,

$$\begin{aligned}\dot{\mathbf{y}}_l^{ST} &= < \dot{\mathbf{y}}_l, \mathbf{t}_l > \\ &= < \dot{\mathbf{s}}_l, \mathbf{t}_l > + \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}(x_{l+u}^{ST} - s_{l+u}^{ST}) < \mathbf{t}_{l+u}, \mathbf{t}_l > + < \dot{\mathbf{v}}_l, \mathbf{t}_l > .\end{aligned} \quad (7.8)$$

Rappelons que :

- $\dot{\mathbf{y}}_l^{ST}$ représente la transformation de la frame attaquées $\dot{\mathbf{y}}_l = [\dot{\mathbf{y}}[Nl], ..., \dot{\mathbf{y}}[Nl + N - 1]] = [\dot{\mathbf{y}}_l[0], ..., \dot{\mathbf{y}}_l[N-1]]$.
- Nous avons, aussi, $\dot{\mathbf{s}}_l = \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}\mathbf{s}_{l+u}$ et $\dot{\mathbf{v}}_l = \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[}\mathbf{v}_{l+u}$ tel que $l = [0, ..., \frac{N\times T}{\tau}]$
- T, N et ω représentent respectivement le nombre de frames vidéo totale, le nombre d'échantillons dans chaque frame et la fenêtre d'attaque.
- Comme nous sommes dans le cas : $\tau = N$ alors : $l = [0, ..., |[T]|]$

Nous avons :

$$\begin{aligned}\dot{\mathbf{y}}_l^{ST} &= <\dot{\mathbf{s}}_l, \mathbf{t}_l> + \frac{1}{\omega}(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) <\mathbf{t}_l, \mathbf{t}_l> + \frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[,u\neq 0}(x_{l+u}^{ST} - s_{l+u}^{ST}) <\mathbf{t}_{l+u}, \mathbf{t}_l> \\ &+ <\dot{\mathbf{v}}_l, \mathbf{t}_l> \end{aligned} \tag{7.9}$$

Le vecteur direction $\mathbf{t}_l$ est considéré normalisé [1] : $<\mathbf{t}_l, \mathbf{t}_l>= 1$, alors :

$$\begin{aligned}\dot{\mathbf{y}}_l^{ST} &= <\dot{\mathbf{s}}_l, \mathbf{t}_l> + \underbrace{\frac{1}{\omega}(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])}_{Informationutile} + \underbrace{\frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[,u\neq 0}(x_{l+u}^{ST} - s_{l+u}^{ST}) <\mathbf{t}_{l+u}, \mathbf{t}_l>}_{Interferences} \\ &+ \underbrace{<\dot{\mathbf{v}}_l, \mathbf{t}_l>}_{projection\ du\ bruit} \end{aligned} \tag{7.10}$$

Cette dernière équation montre deux principaux problèmes :

- L'attaques par moyennage engendre, dans le cas d'un étalement sur une frame, un terme d'interférences :
  $\frac{1}{\omega}\sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[,u\neq 0}(x_{l+u}^{ST} - s_{l+u}^{ST}) <\mathbf{t}_{l+u}, \mathbf{t}_l>$, ce qui perturbe le signal utile (watermark) et diminue donc le w.n.r..
- L'attaques divise la puissance du watermark par $\frac{1}{\omega^2}$, ce qui nous obligera à augmenter le w.n.r. pour compenser les pertes éventuelles et perdre en invisibilité, de plus, pour une large fenêtre d'attaque ω le watermak pourrait être complètement éliminé.

Pour résoudre ces deux problème, nous proposons les deux solutions suivantes :

- Prendre une famille de $\mathbf{t}_l$ orthogonale :

$$\forall i, j \in \{0, ..., \frac{N\times T}{\tau}\} :<\mathbf{t}_i, \mathbf{t}_j>= \delta_i^j$$

  tel que $<,>$ est le produit scalaire entre deux vecteurs et δ_i^j représente le

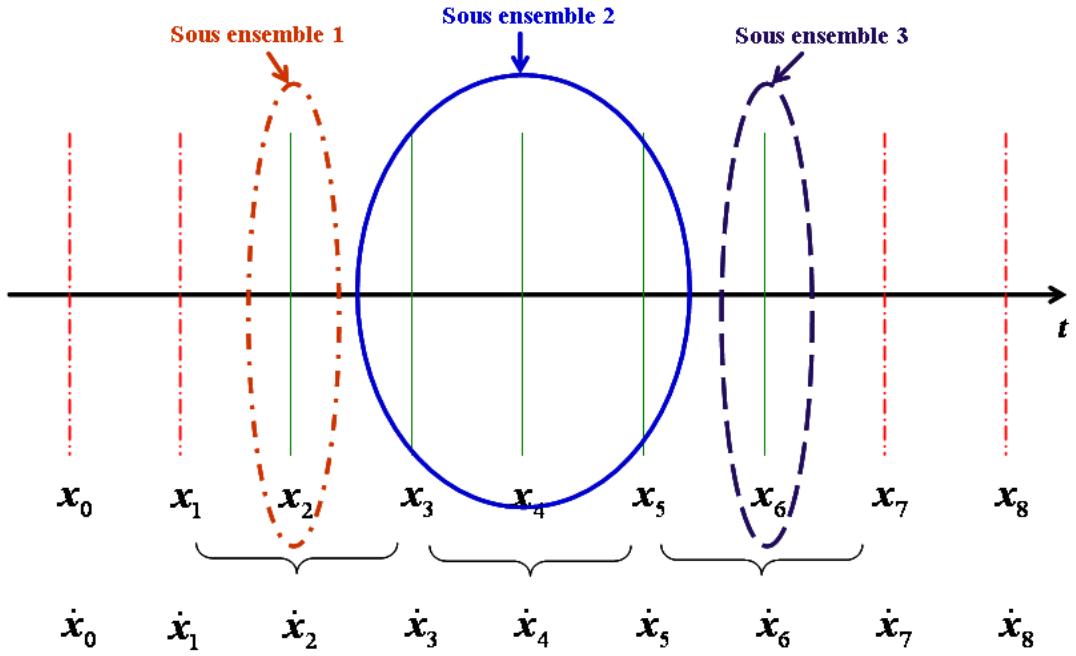


FIGURE 7.2 – Exemple montrant les trois sous ensembles engendrés par une attaque TFA, sur une suite de frames vidéo, avec une fenêtre d'attaque ω égale à 3, où le ST a été utilisé avec un facteur d'étalement sur les frames τ_F égal à 5. Les frames représentées avec des lignes continues contiennent un bit message étalé égal à 1 et celles représentées par des lignes dicontinues continnent un bit message étalé égal à 0.

symbole de Kronecker.

- Étaler le tatouage sur plusieurs frames de manière à ce que l'on puisse réoccuper l'information dans les frames adjacentes utilisées dans l'attaque par moyennage.
- Introduire un nouveau paramètre τ_F qui représente facteur d'étalement sur les frames.

Remarque *Il est très important de faire la différence entre τ_F qui est le facteur d'étalement sur les frames et τ qui représente le facteur d'étalement sur des échantillon.*

7.2 Cas d'un étalement sur plusieurs frames vidéo

Dans ce cas, l'étalement se fait sur τ_F frames, i.e., $\tau = N \cdot \tau_F$, l'insertion du tatouage se fait selon une direction $\mathbf{t'}_l$ qui représente le résultat de la répétition de τ_F fois d'un même vecteurs colonnes $\mathbf{t}_l$: $\mathbf{t'}_l = [t_l, ..., t_l]$ de dimension $N \times 1$ tel que les t_l, $l = [0, ..., T/\tau_F]$ sont orthogonaux entre eux, donc $\mathbf{t'}_l$ aura une dimension $N \cdot \tau_F \times 1$.
Alors la transformation des $N \cdot \tau_F$ échantillons est donné par :

$$\mathbf{s}^{ST}[l] = \sum_{i=N\cdot\tau_F\cdot l}^{N\cdot\tau_F\cdot l+N\cdot\tau_F-1} \mathbf{s}[i] \cdot \mathbf{t}[i], l = [\frac{i}{N \cdot \tau_F}] \tag{7.11}$$

Rappelons que [.] désigne la partie entière d'un nombre réel.
Eqn. peut être simplifiée en transposant la transformation d'échantillon à celle de frames comme suit,

$$\begin{aligned} \mathbf{s}^{ST}[l] &= \sum_{i=N\cdot(\tau_F\cdot l)}^{N\cdot(\tau_F\cdot l)+N-1} \mathbf{s}[i] \cdot \mathbf{t}[i] + \sum_{i=N\cdot(\tau_F\cdot l)+N}^{N\cdot(\tau_F\cdot l)+2N-1} \mathbf{s}[i] \cdot \mathbf{t}[i] \\ &\quad +... + \sum_{i=N\cdot(\tau_F\cdot l)+N\cdot\tau_F-N}^{N\cdot\tau_F\cdot l+N\cdot\tau_F-1} \mathbf{s}[i] \cdot \mathbf{t}[i] \\ &= \sum_{n=\tau_F\cdot l}^{\tau_F\cdot l+\tau_F-1} [\sum_{i=N\cdot n}^{N\cdot n+N-1} \mathbf{s}[i] \cdot \mathbf{t}[i]], \end{aligned} \tag{7.12}$$

alors,

$$\mathbf{s}^{ST}[l] = \sum_{n=\tau_F\cdot l}^{\tau_F\cdot l+\tau_F-1} < \mathbf{s}_n, \mathbf{t}_n > = \sum_{n=\tau_F\cdot l}^{\tau_F\cdot l+\tau_F-1} < \mathbf{s}_n, \mathbf{t}_l >, \tag{7.13}$$

ainsi, la différence entre la transformation de frames et celle d'échantillons réside en l'utilisation du produit scalaire, dans le premier cas, au lieu d'une multiplication.
Il est possible de diviser les frames attaquées en 3 sous ensembles selon les $\ll Interferences \gg$ induites (voir Fig.7.2) engendrées par les frames :

- 1^{er} ensemble : interférences induites par les frames se trouvant en amont de la fenêtre d'étalement,
- 2^{ieme} ensemble : pas d'interférences.
- 3^{ieme} ensemble : interférences induites par les frames se trouvant en aval de la fernêtre d'étalement.

Ainsi, l'attaque TFA a un effet différent pour chaque sous ensemble de frames.

7.2.1 1^{er} ensemble

Il est constitué de tout les frames d'indice n qui vérifient :$n+u < \tau_F l$, sachant que $-\frac{\omega}{2} \leq u < \frac{\omega}{2}$ alors,

$$\tau_F l \leq n \leq \tau_F l + \frac{\omega}{2} - 1. \tag{7.14}$$

La frame attaquée s'écrit dans ce cas :

$$\begin{aligned} \dot{\mathbf{y}}_n &= \underbrace{\frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{s}_{n+u}}_{\dot{\mathbf{s}}_n} + \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \tau_F l - n - 1]} (\mathbf{x}^{ST}[l-1] - \mathbf{s}^{ST}[l-1]) \cdot \mathbf{t}_{n+u} \\ &+ \frac{1}{\omega} \sum_{u \in [\tau_F l - n, \frac{\omega}{2}]} (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_{n+u} + \underbrace{\frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} \mathbf{v}_{n+u}}_{\dot{\mathbf{v}}_n}. \end{aligned} \tag{7.15}$$

7.2.2 2^{ieme} ensemble

Il est composé de frames d'indice n et vérifiant la condition suivante : $\tau_F l \leq n+u \leq \tau_F l + \tau_F - 1$, puisque : $-\frac{\omega}{2} \leq u < \frac{\omega}{2}$, alors

$$\tau_F l + \frac{\omega}{2} \leq n \leq \tau_F l + \tau_F - \frac{\omega}{2}. \tag{7.16}$$

La frame attaquée est formulée donc comme suit :

$$\dot{\mathbf{y}}_n = \dot{\mathbf{s}}_n + \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \frac{\omega}{2}[} (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_{n+u} + \dot{\mathbf{v}}_n \tag{7.17}$$

7.2.3 3^{ieme} ensemble

Dans ce cas, les frames d'indice n constituant ce sous ensemble vérifient la condition : $n+u > \tau_F l + \tau_F - 1 \;\Rightarrow\; \tau_F l + \tau_F - \frac{\omega}{2} + 1 \leq n \leq \tau_F l + \tau_F - 1$ et la frames est donnée alors par l'expression suivante :

$$\dot{\mathbf{y}}_n = \dot{\mathbf{s}}_n + \frac{1}{\omega} \sum_{u \in [-\frac{\omega}{2}, \tau_F l - n - 1]} (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) \cdot \mathbf{t}_{n+u} + \frac{1}{\omega} \sum_{u \in [\tau_F l - n, \frac{\omega}{2}]} (\mathbf{x}^{ST}[l+1] - \mathbf{s}^{ST}[l+1]) \cdot \mathbf{t}_{n+u} + \dot{\mathbf{v}}_n \tag{7.18}$$

Après avoir formulé l'expression des frames attaquée pour les trois ensembles, en utilisant Eqn.7.2 la transformations des frames attaquées est formulée comme suit :

$$
\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \sum_{n=\tau_F l}^{\tau_F l+\tau_F-1} < \dot{\mathbf{y}}_n, \mathbf{t}_n >= \dot{\mathbf{s}}^{ST}[l] \qquad (7.19)\\
&+ \sum_{n=\tau_F l}^{\tau_F l+\frac{\omega}{2}-1} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\tau_F-n-1]} (\mathbf{x}^{ST}[l-1]-\mathbf{s}^{ST}[l-1])\cdot\mathbf{t}_{n+u}, \mathbf{t}_n >\\
&+ < \frac{1}{\omega} \sum_{u\in[\tau_F-n,\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_{n+u}, \mathbf{t}_n >]\\
&+ \sum_{n=\tau_F l+\frac{\omega}{2}}^{n=\tau_F l+\tau_F-\frac{\omega}{2}} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_{n+u}, \mathbf{t}_n >]\\
&+ \sum_{n=\tau_F l+\tau_F-\frac{\omega}{2}+1}^{\tau_F l+\tau_F-1} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\tau_F l-n-1]} (\mathbf{x}^{ST}[l-1]-\mathbf{s}^{ST}[l-1])\cdot\mathbf{t}_{n+u}, \mathbf{t}_n >\\
&+ < \frac{1}{\omega} \sum_{u\in[\tau_F l-n,\frac{\omega}{2}]} (\mathbf{x}^{ST}[l+1]-\mathbf{s}^{ST}[l+1])\cdot\mathbf{t}_{n+u}, \mathbf{t}_n >] + \dot{\mathbf{v}}[l]
\end{aligned}
$$

Sachant que les vecteur direction $\mathbf{t}_n$ sont identiques lorsque $n \in [\tau_F l, \tau_F l + \tau_F + 1]$ et orthogonaux aux reste des vecteur $\mathbf{t}'_n$ tel que n' n'appartient pas à l'intervalle $[\tau_F l, \tau_F l + \tau_F + 1]$, l'equation précedente devient :

$$\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \sum_{n=\tau_F l}^{\tau_F l+\tau_F-1} < \dot{\mathbf{y}}_n, \mathbf{t}_l >= \dot{\mathbf{s}}^{ST}[l] \qquad (7.20)\\
&+ \sum_{n=\tau_F l}^{\tau_F l+\frac{\omega}{2}-1} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\tau_F-n-1]} (\mathbf{x}^{ST}[l-1]-\mathbf{s}^{ST}[l-1])\cdot\mathbf{t}_{l-1}, \mathbf{t}_l >\\
&+ < \frac{1}{\omega} \sum_{u\in[\tau_F-n,\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l >]\\
&+ \sum_{n=\tau_F l+\frac{\omega}{2}}^{\tau_F l+\tau_F-\frac{\omega}{2}} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l >]\\
&+ \sum_{n=\tau_F l+\tau_F-\frac{\omega}{2}+1}^{\tau_F l+\tau_F-1} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\tau_F l-n-1]} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l >\\
&+ < \frac{1}{\omega} \sum_{u\in[\tau_F l-n,\frac{\omega}{2}]} (\mathbf{x}^{ST}[l+1]-\mathbf{s}^{ST}[l+1])\cdot\mathbf{t}_{l+1}, \mathbf{t}_l >] + \dot{\mathbf{v}}[l],
\end{aligned}$$

alors,

$$\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \dot{\mathbf{s}}^{ST}[l] + \sum_{n=\tau_F l}^{\tau_F l+\frac{\omega}{2}-1} < \frac{1}{\omega} \sum_{u\in[\tau_F-n,\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l >] \qquad (7.21)\\
&+ \sum_{n=\tau_F l+\frac{\omega}{2}}^{\tau_F l+\tau_F-\frac{\omega}{2}} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\frac{\omega}{2}[} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l >]\\
&+ \sum_{n=\tau_F l+\tau_F-\frac{\omega}{2}+1}^{\tau_F l+\tau_F-1} [< \frac{1}{\omega} \sum_{u\in[-\frac{\omega}{2},\tau_F l-n-1]} (\mathbf{x}^{ST}[l]-\mathbf{s}^{ST}[l])\cdot\mathbf{t}_l, \mathbf{t}_l > +\dot{\mathbf{v}}[l].
\end{aligned}$$

Comme les $\mathbf{t}_l$ sont normalisés alors,

$$\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \dot{\mathbf{s}}_l^{ST} + \sum_{n=\tau_F l}^{\tau_F l+\frac{\omega}{2}-1} [\frac{1}{\omega\tau_F}((\frac{\omega}{2}) - (\tau_F l - n) + 1)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \\
&+ \sum_{n=\tau_F l+\frac{\omega}{2}}^{\tau_F l+\tau_F-\frac{\omega}{2}} [\frac{1}{\omega\tau_F}((\frac{\omega}{2} - 1) - (-\frac{\omega}{2}) + 1)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \qquad (7.22) \\
&+ \sum_{n=\tau_F l+\tau_F-\frac{\omega}{2}+1}^{\tau_F l+\tau_F-1} [\frac{1}{\omega\tau_F}((\tau_F l + \tau_F - 1) - (-\frac{\omega}{2}) - n + 1)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \\
&+ \dot{\mathbf{v}}[l],
\end{aligned}$$

alors,

$$\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \dot{\mathbf{s}}^{ST}[l] + \sum_{n=\tau_F l}^{\tau_F l+\frac{\omega}{2}-1} [\frac{1}{\omega\tau_F}(\frac{\omega}{2} - \tau_F l + n + 1)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \\
&+ \sum_{n=\tau_F l+\frac{\omega}{2}}^{\tau_F l+\tau_F-\frac{\omega}{2}} [\frac{1}{\omega\tau_F}(\omega)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \\
&+ \sum_{n=\tau_F l+\tau_F-\frac{\omega}{2}+1}^{\tau_F l+\tau_F-1} [\frac{1}{\omega\tau_F}(\tau_F l + \tau_F + \frac{\omega}{2} - n)(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])] \\
&+ \dot{\mathbf{v}}[l] \qquad (7.23)
\end{aligned}$$

Remarque On remarque que dans Eqn.7.23 le premier terme $((\frac{\omega}{2} - \tau_F l + 1) + n \cdot 1)$ ainsi que le deuxième terme : $((\frac{\omega}{2} + \tau_F l + \tau_F) + n \cdot (-1))$ sont de la forme $U_0 + n \cdot r$ qui est aussi la forme d'une suite arithmitique de raison r. Sachant que la somme $Sum = U_p + ... + U_n$, $p \in \mathcal{N}$ et $n \in \mathcal{N}$ est donnée par la formule suivante :

$$Sum = \frac{(n - p + 1)(U_n + U_p)}{2}$$

d'où

$$\begin{aligned}
\dot{\mathbf{y}}^{ST}[l] &= \dot{\mathbf{s}}^{ST}[l] \\
&+ [\frac{1}{\omega\tau_F}(\tau_F l + \frac{\omega}{2} - 1 - \tau_F l + 1)\frac{(\frac{\omega}{2} - \tau_F l + \tau_F l + \frac{\omega}{2} - \tau_F l + \tau_F l + \frac{\omega}{2} + 1)}{2}] \\
&+ [(\tau_F l + \tau_F - \frac{\omega}{2} - \tau_F l - \frac{\omega}{2} + 1)\frac{1}{\tau_F}] \\
&+ [(\tau_F l + \tau_F - 1 - [(\tau_F l + \tau_F - \frac{\omega}{2} + 1]) + 1) \\
&\cdot \ (\frac{\tau_F l + \tau_F + \frac{\omega}{2} - (\tau_F l + \tau_F - 1) + \tau_F l + \tau_F + \frac{\omega}{2} - (\tau_F l + \tau_F - \frac{\omega}{2} + 1)}{2\omega\tau_F}) \\
&\cdot \ (\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l])]) \qquad (7.24) \\
&+ \dot{\mathbf{v}}^{ST}[l]
\end{aligned}$$

$$\dot{\mathbf{y}}_l^{ST} = \dot{s}_l^{ST} + (\frac{\frac{3}{4}\omega + 1}{4\tau_F} + \frac{\tau_F - \omega + 1}{\tau_F} + \frac{(\frac{3\omega}{4} - \frac{3}{2})}{2\tau_F})(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) + \dot{\mathbf{v}}[l] \qquad (7.25)$$

Finalement, la frame attaqué peut être exprimée par :

$$\dot{\mathbf{y}}^{ST}[l] = \mathbf{s}^{ST}[l] + (\frac{4\tau_F - \frac{7}{4}\omega + 2}{4\tau_F})(\mathbf{x}^{ST}[l] - \mathbf{s}^{ST}[l]) + \dot{\mathbf{v}}^{ST}[l] \qquad (7.26)$$

Chapitre 8

Appendix C
Liste des publications

Ci-dessous la liste des articles publiés/soumis dans les revues et conférences internationales et nationales :

1. Braci S., Delpha C. and Boyer R., Informed Stego-schemes in Active Warden Context : Tradeoff between steganographic performance, submitted in Elsevier Journal of Signal Processing : Image Communication.
2. Braci S., Delpha C. and Boyer R., How Quantization Based Schemes can be Used in Steganographic Context submitted in Elsevier Journal of Signal Processing : Image Communication.
3. Braci S., Boyer R. and Delpha C., Analysis of the Resistance of the Spread Transform Against Temporal Frame Averaging attack, International Conference on Image Processing (ICIP), Hong Kong, China, September, 2009.
4. Braci S., Boyer R. and Delpha C., Security evaluation of informed watermarking scheme, International Conference on Image Processing (ICIP), Cairo, Egypt, November, 2009.
5. Benkara Mostefa I., Braci S., Delpha C., Boyer, R. et Khamadja M., Etude du schéma Scalaire de Costa dans un domaine indépendant, GRETSI, Dijon, France, Septembre, 2009.
6. Benkara Mostefa I., Braci S., Delpha C., Boyer R., and Khamadja M., Improved Performances of Scalar Costa Scheme for Images Watermarking in an Independent Domain, 6th Int'l Symposium on Image and Signal Processing and Analysis, Salzburg, Austria, September, 2009.

7. Braci S., Delpha C. and Boyer R., How quantization based scheme can be used in steganographic context, International Workshop on Multimedia Signal Processing (MMSP), Rio de Janeiro, Brazil, October, 2009.
8. Braci S., Miraoui A., Delpha C. and Boyer R., Watermarking scar as an ultimate copy protection, Euro American Workshop on Information Optics (WIO), July, 2009.
9. Maity S.P., Delpha C., Braci S. and Boyer R., Hidden QIM Watermarking on Compressed Data using Channel Coding and Lifting, December, International Conference on Pattern Recognition and Machine Intelligence PREMI-09, New Delhi, India, 2009.
10. Braci S., Delpha C., Boyer R. and Le Guelvouit, G., Informed stego-systems in active warden context : Statistical undetectability and capacity, IEEE International Workshop on MultiMedia Signal Processing (MMSP), Cairns, Australia, October, 2008.
11. Braci S., Boyer R. and Delpha C., On the tradeoff between security and robustness of the Trellis Coded Quantization scheme, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), April, 2008.

Références bibliographiques

[1] J. J. Eggers, R. Bauml, R. Tzchoppe, and B. Girod. Scalar costa scheme for information embedding. *IEEE Trans. on Signal Processing*, 51 :1003–1019, 2003.

[2] P. Guillon, T. Furon, and P. Duhamel. Applied public-key steganography. In *Proc. SPIE*, volume 3710, pages 38–49, 2002.

[3] F. A. P. Petitcolas, R. J. Anderson, and M. G. Kuhn. Information hiding : a survey. In *Proceedings of the IEEE, special issue on protection of multimedia content*, volume 87, pages 1062–1078, july 1999.

[4] G. Doërr. *Security Issue and Collusion Attacks in Video Watermarking*. phd thesis, Université de Nice-Sophia Antipois, 2005.

[5] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker. *Digital watermarking and steganography*. Morgan Kaufmann, 2008.

[6] Herodotus. *The histories*. Penguin books, 1996.

[7] F.Y Shih. *Digital watermarking and steganography*. CRC Press, Taylor & Francis Group, 2008.

[8] M. Barni and F. Bartolini. *Watermarking systems engineering*. Signal processing and communications series, 2004.

[9] B. Furht and D. Kirovski. *Multimedia watermarking techniques and applications*. Auberbach publications, Taylor & Francis Group, 2006.

[10] K.J.R. Liu, W. Trappe, Z.J. Wang, M. Wu, and H. Zhao. *Multimedia Fingerprinting Forensics for Traitor Tracing*. EURASIP Book Series on Signal Processing and Communications, Hindawi Publishing Co., 2005.

[11] G. Coatrieux, B. Sankur, and H. Maitre. Strict integrity control of biomedical images. *Proceeding SPIE*, 4314 :229–240, 2001.

[12] M. M. Yeung and F. C. Mintzer. Invisible watermarking for image verification. *Journal of electronic imaging*, 7 :578–591, July 1998.

[13] F. Hertung, I. Cox, T. Kalker, J.-P. Linnartz, M. Miller, and C. Traw. Digital watermarking of raw and compressed video, digital compression systems for video communication. 66 :205–213, 1996.

[14] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. on Information Theory*, 29 :439–441, 1983.

[15] B. Chen and G. W. Wornell. Quantization index modulation : a class of provably good methods for digital watermarking and information embedding. *IEEE Trans. on Information Theory*, 47 :1423–1443, 2001.

[16] P. Moulin and R. Koetter. Data-hiding codes. *(tutorial paper),IEEE Trans. Information Theory, special issue on Security*, 93 :2783–2127, 2005.

[17] J. R. Hernández O. Pérez-gonzález. A tutorial on digital watermarking. In *IEEE Annual Carnahan Conference on Security Technology*, october 1999.

[18] D. J. Granrath. The role of human visual models in image processing. In *Proce. IEEE*, volume 69, may 1981.

[19] D. J. M. Robinson and M. J. Hawksford. The role of human visual models in image processing. In *Proce. IEEE*, volume 69, may 1981.

[20] D. Coltuc and J-M. Chassery. Mapping based reversible watermarking. In *4th International Conference on Sciences of Electronic, Technologies of Information and Telecommunications*, pages 70000V :1–10, Hammamet, Tunisie.

[21] M. Chaumont and W. Puech. A high capacity reversible watermarking scheme. In *Electronic Imaging, Visual Communications and Image Processing*, San Jose, USA, january 2009.

[22] A.M. Alattar. Reversible watermark using the difference expansion of a generalized integer transform. *IEEE Trans. on Image Processing*, 13 :1147–1156, 2004.

[23] W. Puech, J.M. Rodrigues, and J.E. Develay-Morice. A new fast reversible method for image safe transfer. *Journal of Real-Time Image Processing, Springer*, 2 :55–65.

[24] W. Puech K. Hayat and G. Gesquière. Scalable 3d terrain visualization through reversible jpeg2000-based blind data hiding. *IEEE Trans. on Multimedia*, 10 :1261–1276.

[25] F. Cayre, C. Fontaine, and T. Furon. Watermarking security : Theory and practice. *IEEE Transactions on Signal Processing*, 53(10) :3976–3987, 2005. numéro spécial "Supplement on Secure Media III".

[26] I.J. Cox, J. Kiliany, T. Leightonz, and T. Shamoony. Secure spread spectrum watermarking for multimedia. *IEEE Trans. on Image Processing*, 6 :1673–1687, december 1997.

[27] C.E. Shannon. Channels with side information at the transmitter. *IBM Journal of Research and Development*, 2 :289–293, 1958.

[28] S. I. Gel'fand and M. S. Pinsker. Problems of control theory. 9 :19–31, 1980.

[29] C. Heegard and A. A. El Gamal. On the capacity of computer memory with defects. *IEEE Trans. on Information Theory*, 29 :731–739, 1983.

[30] M.W. Marcellin and T.R. Fischer. Trellis-coded quantization of memoryless and gauss-markov sources. *IEEE Trans. on Communication*, 38 :83–93, 1990.

[31] G. Ungerboeck. Channel coding with multilevel/phase signals. *IEEE Trans. on Information Theory*, 28 :55–67, 1982.

[32] A. J. Viterbi and J. K. Omura. *Principle of digital communication and coding.* McGraw-Hill book company, 1979.

[33] E. Esen, A.A. Alatan, and M. Aska. Trellis coded quantization for data hiding. In *EUROCON*, pages 24–24, 2003.

[34] G. J. Simmons. The prisonners' problem and the subliminal channel. In *Advances in Cryptology : Proceedings of CRYPTO*, pages 51–67. Plenum Press, 1984.

[35] J. Jacod and A.N Shiryaev. *Limit Theorems for Stochastic Processes.* Springer-Verlag Berlin, 2002.

[36] J. E. Vila-forcén, S. Voloshynovskiy, O. Koval, F. Pérez-gonzález, and T. Pun. Practical datahiding : Additive attacks performance analysis. In *In International Workshop on Digital Watermarking*, volume 3710, pages 244–259. Springer Verlag, 2005.

[37] Y. Wang and P. Moulin. Perfectly secure steganography : capacity, error exponents, and code constructions. *IEEE Trans. Information Theory, special issue on Security*, pages 2706–2722, June 2008.

[38] D. Kundur and D. Hatzinakos. A robust digital image watermarking method using wavelet-based fusion. In *Proceedings on International Conferance on Image Processing (ICIP)*, volume 1, pages 544–547, 1997.

[39] N. Cvejic and T. Seppnen. Improving audio watermarking scheme using psychoacoustic watermark filtering. In *Proc. of the first IEEE International Symposium on Signal Processing and IT*, pages 163–168, 2001.

[40] Ying Wang and Pierre Moulin. Steganalysis of block-structured stegotext. In *Security, Steganography, and Watermarking of Multimedia Contents*, volume 5306, pages 477–488, 2004.

[41] A. Gersho and R. M. Gray. *Vector quantization and signal compression*. Kluwer academic publishers, 1992.

[42] Joachim J. Eggers and R. Biuml. A communications approach to image steganography. In *Proc. SPIE*, volume 4675, pages 26–37, 2002.

[43] T. Cover and J. Thomas. *Elements of Information Theory*. Wiley, 1991.

[44] S. I. Gel'fand and M. S. Pinsker. Coding for channel with random parameters. 9 :19–31, 1980.

[45] R. J. Anderson and F. A. P. Petitcolas. On the limits of steganography. *IEEE Journal of Selected Areas in Communication*, 16 :474–481, 1998.

[46] G. Le Guelvouit. Trellis-coded quantization for public-key steganography. In *accepted to IEEE Conf. on Acoustics, Speech and Signal Proc.*, march 2005.

[47] S. Braci, R. Boyer, and C. Delpha. Security evaluation of informed watermarking scheme. In *International Conference on Image Processing (ICIP)*, November 2009.

[48] S. Braci, A. Miraoui, C. Delpha, and R. Boyer. Watermarking scar as an ultimate copy protection. In *Euro American Workshop on Information Optics (WIO)*, July 2009.

[49] S. Braci, R. Boyer, and C. Delpha. Analysis of the resistance of the spread transform against temporal frame averaging attack. In *International Conference on Image Processing (ICIP)*, September 2009.

[50] K. Tanaka, Y. Nakamura, and K. Matsui. Embedding secret information into a dithered multi-level image. In *In Proc. of Military Communications. Conference*. IEEE, 1990.

[51] A.Z. Tirkel, G.A. Rankin, R.M. Van Schyndel, W.J. Ho, N.R.A. Mee, and C.F. Osborne. Electronic watermark. In *DICTA conference*, pages 666–673, Sydney, Australia.

[52] S. Braci, R. Boyer, and C. Delpha. On the tradeoff between security and robustness of the trellis coded quantization scheme. In *IEEE International Conference on Accoustics, Speech and Signal Processing (ICASSP)*, pages 1733–1736, april 2008.

[53] M. Barni and F. Bartoloni. *Watermarking systems engineering.* Signal processing and communication series, 2004.

[54] F. Cayre, C. Fontaine, and T. Furon. Watermarking security : Theory and practice. *IEEE Trans. on Signal Processing*, 53 :3976–3987, 2005.

[55] A. Kerckhoffs. *La cryptographie militaire.* Journal des sciences militaires, 1983.

[56] M. Barni, F. Bartolini, and T. Furon. A general frameworkfor robust watermarking security. *Signal processing revue, Elsivier*, 83 :2069–2084, 2003.

[57] R.G. Gallager. *Information theory and reliable communication.* John WILLEY & SONS, 1969.

[58] J. G. Proakis. *Digital communications.* McGraw-Hill book company, 2000.

[59] C.E. Shannon. Communication theory of secrecy systems. *Bell System Technical Journal*, 28 :656–715, 1949.

[60] S. Voloshynovskiy L. Pérez-Freire, F. Pérez-González. A security risk for publicly available watermark detectors. In *Benelux Information Theory Symposium*, veldhoven, The Netherlands, 1998.

[61] J-L. Dugelay G. Doerr. Security pitfalls of frame-by-frame approaches to cideo watermarking. *IEEE Trans. on signal processing*, 52(7) :2955–2964, 2004.

[62] J-L Dugelay G. Doerr. New intra-video collusion attack using mosaicing. In *IEEE proceeding on multimedia and expo (ICME)*, pages 505–508, July 2003.

[63] R. Caldelli, A. Piva, M. Barni, and A. Carboni. Effectiveness of st-dm watermarking against intra-video collusion. *Lecture Notes in Computer Science*, 3710 :158–170, 2005.

[64] G. Doërr. *Security Issue and Collusion Attacks in Video Watermarking.* Thse de doctorat, Université de Nice-Sophia Antipois, 2005.

[65] E. Scott. *Computer vision and image processing.* Prentice Hall PTR, 1999.

[66] R. Rao and S. Dianat. *Basics of Code Division Multiple Access (CDMA).* SPIE Press Book, 2005.

[67] S. K. Pal. Fast, reliable and secure digital communication using hadamard matrices. In *IEEE Computer Society, Proceedings of the International Conference on Computing : Theory and Applications.*

[68] J. J. Sylvester. Thoughts on inverse orthogonal matrices, simultaneous sign successions, and tessellated pavements in two or more colours, with applications to newton's rule, ornamental tile-work, and the theory of numbers. *Philosophical Magazine*, 34 :461–475, 1867.

[69] C. Cachin. *An information-theoretic model for steganography.* Information Hiding, 1998.

[70] S. Braci, C. Delpha, R. Boyer, and G. Le Guelvouit. Informed stego-systems in active warden context : Statistical undetectability and capacity. In *IEEE Proceedings Multimedia Signal Processing (MMSP)*, pages 707–712, october 2008.

[71] Z. Shahid, M. Chaumont, and W. Puech. Fast protection of h.264/avc by selective encryption. In *Singaporean-French IPAL Symposium*, volume 9, pages 18–20, Fusionopolis, Singapore, february.

[72] S. Duta, M. Mitrea, M. Belhaj, and F. Preteux. A comparative study on insertion strategies in mpeg 4 avc watermarking. In *Proceedings SPIE Conference on Mathematics of Data/Image Pattern Recognition, Compression, and Encryption with Applications XI*, volume 7075, page 707509, San Diego, CA, august 2008.

[73] S. Duta, M. Mitrea, and F. Preteux. Compressed versus uncompressed video watermarking. In *Proceedings SPIE Conference on Mathematics of Data/Image Pattern Recognition, Compression, Coding, and Encryption with Applications X*, volume 6700, pages 67000A :1–12, San Diego, CA, august 2007.

[74] http ://www.encoding.com/.

[75] ITU-T. *Video codec for audiovisual services at px64 kbits/s.* Technical report, ITU-T Rec. H.120, 1984.

[76] A. Habibi. Hybrid coding of pictorial data. *IEEE Trans. on Communications*, 1974.

[77] ITU-T. *Video codec for audiovisual services at px64 kbits/s.*

[78] ISO/IEC JTC 1. *Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbis/s - part 2 : Video.* Technical report, ISO/IEC 11172-2 (MPEG-1), 1993.

[79] ISO/IEC JTC 1/SC 29. *Generic coding of moving pictures and associated audio infor-mation : Systems.* Technical report, ISO/IEC 13818-1 (MPEG-2 Part 1), 1996.

[80] ISO/IEC JTC 1/SC 29. *Video coding for low bit rate communication.* ITU-T Rec.H.263, 1995.

[81] ISO/IEC JTC 1. *Coding of audio-visual objects - part 2 : Video.* Technical report, ISO/IEC 14496-2 (MPEG-4 visual version 1), 1999.

[82] ISO/IEC JTC 1. *Advanced video coding for generic audiovisual services.* Technical report, ITU-T Rec. H.264, and ISO/IEC 14496-10 AVC, 2003.

[83] G. Bjontegaard and K. Lillevold. *Context-adaptive VLC coding of coefficients.* Technical report, JVT,, 2002.

[84] T. Weigand D. Marpe, H. Schwarz. Context-based adaptive binary arithmetic coding in the h.264/avc video compression standard. *IEEE Trans. on Circuits and Systems for Video Technology*, 13(7) :620–636, 2003.

[85] D. Marpe, G. Blattermann, and T. Wiegand. *Adaptive codes for H.26L. Technical repor.* Technical report, JVT, 2001, 2001.

[86] H.S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky. Low-complexity transform and quantization in h.264/avc. *IEEE Trans. on Circuits and Systems for Video Technology*, 13 :598–603, 2003.